



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

3.38 Protein Crystallography in Drug Discovery

T Hogg and R Hilgenfeld, University of Lübeck, Lübeck, Germany

© 2007 Published by Elsevier Ltd.

3.38.1	Introduction: The Cost and Value of Macromolecular Crystallography in Drug Discovery	875
3.38.2	Macromolecular Crystallography and Rational Drug Design: A Historical Perspective	876
3.38.3	Crystallographic Milestones in Rational Drug Design	877
3.38.3.1	Milestone 1: Gene to Protein	877
3.38.3.1.1	Domain mapping and rational deoxyribonucleic acid construct design	879
3.38.3.1.2	Affinity chromatography: finding the right tag	880
3.38.3.1.3	Choosing a suitable expression system	880
3.38.3.2	Milestone 2: From Protein to Crystal	880
3.38.3.2.1	Solubility optimization: the first hurdle between purification and crystallization	881
3.38.3.2.2	Protein characterization: fit for crystallization?	881
3.38.3.2.3	Dealing with heterogeneity	881
3.38.3.2.4	The protein as a variable	884
3.38.3.3	Milestone 3: From Crystal to Structure	888
3.38.3.3.1	Postcrystallization treatments to improve diffraction quality	888
3.38.3.3.2	Cryocrystallography in practice	889
3.38.3.3.3	Phasing methods for drug discovery	889
3.38.3.4	Milestone 4: Structure to Lead	890
3.38.3.4.1	Crystallography-driven lead characterization: the traditional view	890
3.38.3.4.2	Crystallography and 'fragonomics': the emerging area of fragment-based drug design	892
3.38.4	The Era of High-Throughput Crystallography in Lead Discovery	894
	References	894

3.38.1 Introduction: The Cost and Value of Macromolecular Crystallography in Drug Discovery

Drug discovery has traditionally relied either on serendipitous observations (a classic example would be Sir Alexander Fleming's discovery of the antibacterial properties of penicillin¹), or on screening of natural and synthetic compounds in combination with medicinal chemistry (*see* 1.07 Overview of Sources of New Drugs; 3.32 High-Throughput and High-Content Screening). While remaining indispensable to the drug discovery process, the conventional methods are beginning to give way to more rational approaches that utilize available knowledge on the three-dimensional (3D) structure of a drug target (*see* 4.24 Structure-Based Drug Design – The Use of Protein Structure in Drug Discovery).² This information can be generated experimentally by employing biophysical methods such as x-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy, or theoretically by utilizing experimentally derived 3D structures of related macromolecules to generate 3D homology models of the drug target (*see* 3.21 Protein Crystallography; 3.22 Bio-Nuclear Magnetic Resonance; 4.10 Comparative Modeling of Drug Target Proteins).³

The method of x-ray crystallography has been the most productive generator of experimentally derived 3D structures of biological macromolecules. Of the approximately 14 000 nonredundant macromolecular structures (<95% sequence identity) deposited to the Protein Data Bank (PDB) at the time of this writing, more than 70% have been determined by x-ray crystallography, eclipsing all other biophysical methods combined (e.g., NMR spectroscopy, cryo-electron microscopy, neutron diffraction, and electron diffraction) (*see* 3.17 The Research Collaboratory for Structural Bioinformatics Protein Data Bank). Owing to a great number of technological advances in the field of x-ray crystallography in recent years (many of which will be reviewed in this chapter), the method is certain to maintain its dominant position in structure-based drug design (SBDD) (*see* 4.24 Structure-Based Drug Design – The Use of Protein

Structure in Drug Discovery). Recent estimates have underscored the value of an x-ray crystal structure in the drug discovery process: the average cost of \$15–20 million required for a successful round of lead identification to investigational new drug filing can be reduced by an estimated 50% if 3D structure is centrally utilized in the discovery process.⁴ The largest cost reductions stem from the increased quality of lead candidates and number of different pharmacophore series that can be designed with the aid of structural information. Considering the great value of a 3D structure, the relative cost of an x-ray structure determination is nearly negligible, with the elucidation of novel soluble targets averaging between \$140 000 (bacterial) and \$450 000 (human). On the other hand, the successful x-ray crystal structure determination of integral membrane proteins, which represent the largest class of human drug targets and are particularly recalcitrant to crystallization, can cost millions of dollars (*see* 2.19 Diversity versus Focus in Choosing Targets and Therapeutic Areas).⁴ Ongoing developments in the field of soluble and membrane protein crystallization are poised to reduce further the attrition rates and average cost of x-ray structure determination in the coming years.

3.38.2 Macromolecular Crystallography and Rational Drug Design: A Historical Perspective

The utility of structural models in the rational design of therapeutically useful molecules had its humble beginnings in the mid-1970s, when two researchers at The Squibb Institute of Medical Research, David Cushman and Miguel Ondetti, attempted to design small molecule inhibitors of angiotensin-converting enzyme (ACE), a key enzyme involved in the regulation of blood pressure. ACE, a dipeptidyl carboxypeptidase, cleaves the C-terminal His-Leu dipeptide from the prohormone angiotensin I (Asp-Arg-Val-Tyr-Ile-His-Pro-Phe-His-Leu) to produce angiotensin II, a potent vasopressor octapeptide. Back then, only a handful of macromolecular x-ray crystal structures had been elucidated, one being carboxypeptidase A, a Zn^{2+} -containing exopeptidase, which had been a research focus of William Lipscomb's group at Harvard.^{5–7} Although the structure of ACE was not known at the time, biochemical evidence suggested the two enzymes might be similar: both could be inhibited by metal-chelating agents such as EDTA, both could be reactivated by Zn^{2+} , Mn^{2+} , and Co^{2+} , and both were inactive against substrate analogs containing terminal carboxamide or dicarboxylic amino acids. The working hypothesis that Cushman and Ondetti had formulated was that ACE was a Zn^{2+} -metallopeptidase bearing an active site similar to carboxypeptidase A, with ACE cleaving dipeptides rather than single amino acids from the C-terminus of peptide substrates.

Guided by Lipscomb's structures^{5–7} and published biochemical data reporting inhibition of carboxypeptidase A by D-benzylsuccinic acid,⁸ Cushman and Ondetti sketched out a model of substrate and inhibitor binding to the ACE active site (**Figure 1**). In creating their model, they postulated that the potent inhibitory properties of D-benzylsuccinic acid against carboxypeptidase A ($K_i = 4.5 \times 10^{-7}$ M) originated from a specific binding mode that optimized interactions between the inhibitor and the important substrate-binding groups at the enzyme's active site. Several rationalizations were included in the design process:

1. ACE, like carboxypeptidase A, probably carries a positively charged group, which interacts with the carboxy-terminal carboxylate moiety of the ligand.
2. The active-site Zn^{2+} of ACE and carboxypeptidase A is suitably located to polarize the carbonyl group of the substrate's scissile peptide bond. The metal is likely coordinated by the anionic succinate moiety of D-benzylsuccinic acid in the case of carboxypeptidase A.
3. By virtue of ACE's dipeptidyl carboxypeptidase activity, the distance between its positively charged carboxyl-binding site and its active-site Zn^{2+} should be greater than the corresponding distance in carboxypeptidase A by the length of approximately one amino acid. An additional 'spacer' would therefore need to be incorporated into an ACE inhibitor.
4. All naturally occurring peptidic ACE inhibitors, such as teprotide (isolated from the venom of the Brazilian pit viper), carry a proline at the C-terminus, suggesting that this feature is critical and should be preserved in nonpeptidic inhibitors as well.

Armed with their simple pencil-sketch, Cushman and Ondetti set out to create potent nonpeptidic ACE inhibitors. Their first compound, succinyl-L-proline, indeed proved to be a specific inhibitor of ACE; however, it was only slightly active. By exploring different structural modifications, a simple methylation of the succinyl moiety was found to increase inhibitory potency by more than one order of magnitude. Further attempts to improve the binding affinity of this compound, 2-D-methylsuccinyl-L-proline, steered their decision to explore the role of the inhibitor's probable Zn^{2+} -binding moiety. Replacement of the succinyl carboxyl group with a sulfhydryl function (**Figure 1**) reduced the 50% inhibition constant (IC_{50}) by an additional three orders of magnitude.⁹ The resulting

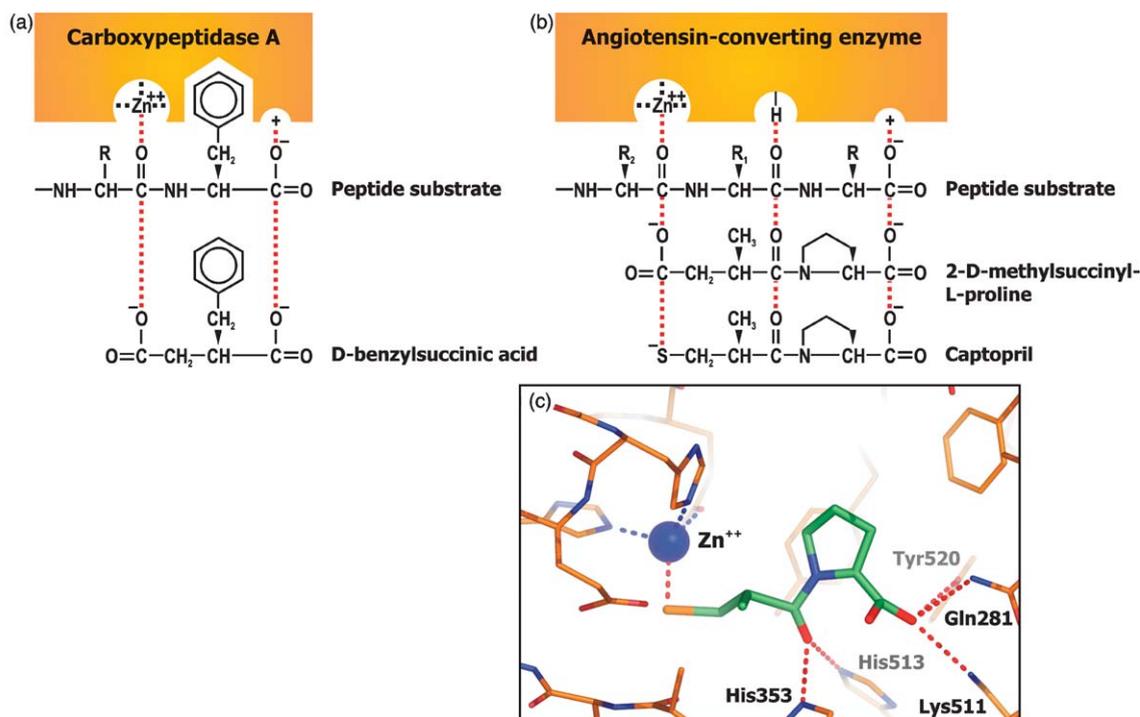


Figure 1 Structure-based design of captopril. (a) Schematic representation of substrate and inhibitor binding to the active site of pancreatic carboxypeptidase A derived from crystallographic data. (b) Cushman and Ondetti's hypothetical model⁹ of the active site of ACE and proposed modes of substrate and inhibitor binding, leading to the design of captopril in 1977. Hydrogen bonds and ionic interactions are represented by dashed lines. (c) Confirmation of Cushman and Ondetti's model arising from the crystal structure elucidation of human testes ACE in complex with captopril in 2004.¹¹ Captopril is depicted in green. (Panels (a) and (b) are reprinted with permission from Ondetti, M. A.; Rubin, B.; Cushman, D. W. *Science* **1977**, *196*, 441–444. Copyright 1977 AAAS.)

compound, 2-D-methyl-3-mercaptopropanoyl-L-proline ('captopril', marketed under the tradename Capoten), turned out to be a potent oral ACE-specific antihypertensive, received FDA approval in the early 1980s, and became Squibb's first billion-dollar drug. Nearly three decades later, the recent crystal structure determinations of ACE:inhibitor complexes^{10,11} finally confirm that – despite ACE bearing little overall structural similarity to carboxypeptidase A – the original pencil-and-paper models of ACE-inhibitor binding were remarkably accurate (Figure 1). Cushman and Ondetti's seminal work represents a landmark in the field of SBDD.

3.38.3 Crystallographic Milestones in Rational Drug Design

The principle stages of an x-ray crystallography project within an SBDD program are typically represented by the following milestones: (1) satisfactory overexpression, purification, and solubilization of the target macromolecule; (2) reproducible crystallization of the target in a form suitable for high-resolution x-ray data analysis; (3) collection of high-quality x-ray diffraction data (unliganded target and/or target:ligand complexes), successful phase determination, and production of a high-quality 3D structural model; and (4) crystallography-driven lead discovery and/or lead optimization (*see* 3.19 Protein Production for Three-Dimensional Structural Analysis; 3.20 Protein Crystallization; 3.21 Protein Crystallography; 3.23 Protein Three-Dimensional Structure Validation; 3.24 Problems of Protein Three-Dimensional Structures).

3.38.3.1 Milestone 1: Gene to Protein

One of the most challenging aspects of any crystallography project is the design of initial gene constructs coding for the protein of interest. Subsequent gene expression as well as purification and characterization of the recombinant gene product can also represent one of the most time-consuming stages in an SBDD program (Figure 2). Milligram quantities

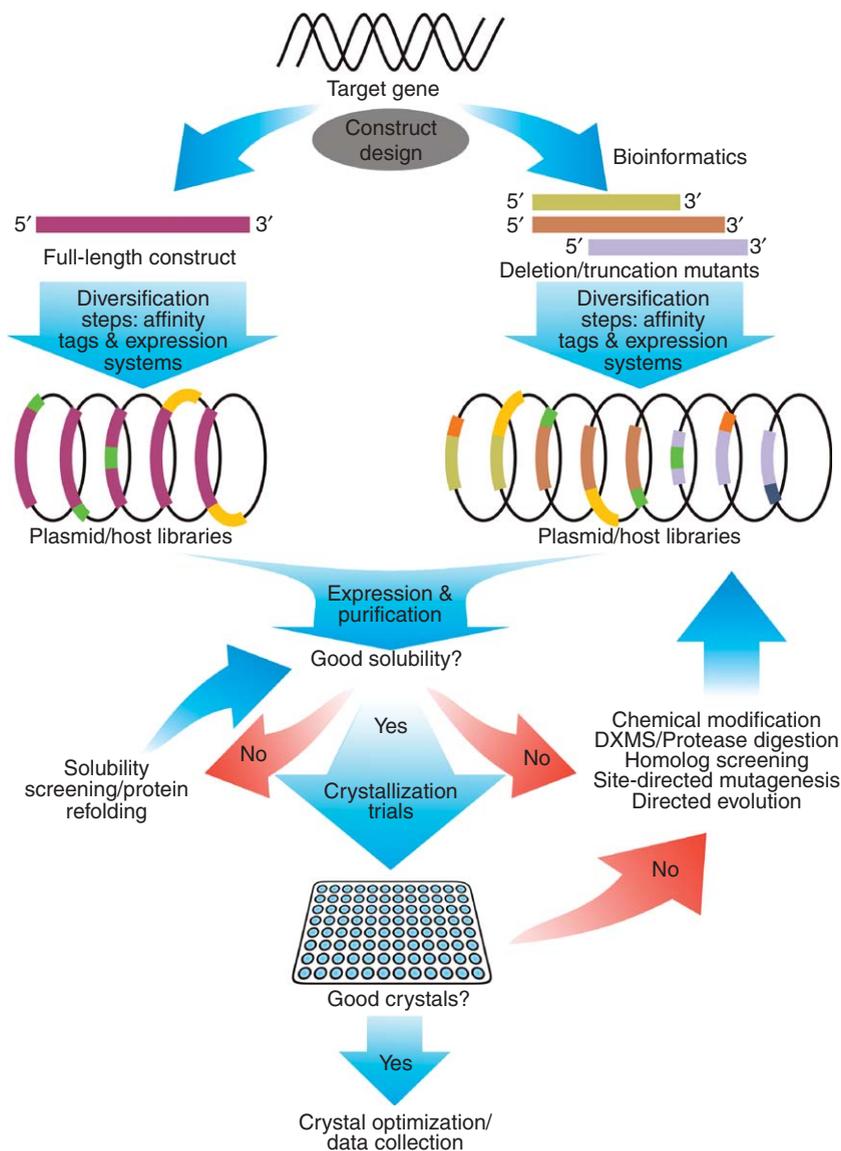


Figure 2 Flow chart illustrating the typical gene-to-crystal bottlenecks and milestones of an SBDD program. Because the probability of success is exponentially related to the number of independent protein-target variants screened, diversification should be emphasized at the stages of gene construct design, choice of expression system, and selection of affinity tags. Available biochemical, biophysical, and bioinformatical data should be exploited during the design of deletion/truncation mutants. Achievement of the first major milestone is commonly regarded as the obtaining of pure, homogeneous, monodisperse, and highly soluble samples of one or more recombinant proteins. Samples that fail to meet quality criteria can often be rescued by implementing feedback loops that incorporate different screening and optimization procedures such as high-throughput solubility screening. If extensive crystallization screening fails to produce initial crystal ‘hits,’ contingency pathways should be followed that involve modification protocols such as site-directed mutagenesis or directed evolution.

of pure protein (typically 0.5–5.0 mg) are routinely needed to screen a sufficient number of crystallization conditions in order to generate initial crystal ‘hits.’ In the past, a small number of deoxyribonucleic acid (DNA) constructs would be generated, subsequently tested for expression in the host cell, and the levels of soluble recombinant protein produced would be assayed by polyacrylamide gel electrophoresis (PAGE). This process was conducted sequentially for each individual gene construct, and the best were used to produce protein for downstream crystallization trials. Recent technological advances have now made it possible to rapidly generate and test multiple gene constructs in parallel.^{12–14}

3.38.3.1.1 Domain mapping and rational deoxyribonucleic acid construct design

Because of the vast amount of uncertainties that go into designing initial gene constructs, structural information should be used whenever possible in the design process. Available 3D structures of homologs, orthologs, or paralogs are a rich source of information regarding domain boundaries and unstructured loop regions, which need to be considered when designing crystallizable targets (*see* 3.17 The Research Collaboratory for Structural Bioinformatics Protein Data Bank). Because multidomain proteins can exhibit a high degree of conformational flexibility, which might preclude crystallization, trimming down the target to a more compact form consisting of the catalytic core represents a logical strategy, particularly for the purposes of SBDD. Some recent examples pertaining to this approach include studies on the kinase domain of c-Abl,¹⁵ and our own work on the bifunctional catalytic domain of the bacterial stringent response factor, RelA/SpoT.¹⁶ Proteolytic mapping of the full-length protein with subsequent analysis by SDS-PAGE can be a useful technique for identifying stable domains. Typically, the full-length protein is digested with a panel of proteases, and samples are analyzed either at fixed time points with different protease concentrations, or at variable time points with fixed protease concentrations. The protein can also be digested in the presence of known ligands or inhibitors, since ligand-induced ordering of unstructured regions may lead to unique digestion maps and suggest alternative possibilities for DNA construct design. Characterization of stable fragments by N-terminal sequencing or mass spectrometry is subsequently carried out to provide a framework for gene construct optimization.¹⁷

Recent years have seen an explosion of useful bioinformatics tools, which can aid in gene construct design by predicting disordered regions in a given protein sequence (*see* 3.15 Bioinformatics; 3.16 Gene and Protein Sequence Databases). These include FoldIndex,¹⁸ DisEMBL,¹⁹ DISOPRED2,²⁰ DRIP-PRED (R. M. MacCallum, unpublished data), GlobPlot 2,²¹ IUPred,²² PONDR,²³ Prelink,²⁴ RONN,²⁵ and the VL2/VL3/VL3H/VL3E suite,^{26,27} all of which utilize different methods (e.g., neural networks, support vector machines) for disorder prediction. The bi-annual CASP (Critical Assessment of Techniques for Protein Structure Prediction) challenge, which constantly evaluates and ranks the accuracy of the various structure prediction programs, has noted very significant progress in the field.²⁸ A curated database containing information on proteins with partial or complete disorder, DisProt, is under development.²⁹

Perhaps one of the most exciting recent developments that promises to greatly assist the process of gene construct design is deuterium-exchange mass spectrometry (DXMS) (Figure 3).^{30,31} This method, which produces high-resolution maps of ordered and disordered regions along the protein sequence, requires only micrograms of soluble protein, can detect disordered segments as little as four residues, can be used in combination with ligands and inhibitors, and has demonstrated success in producing crystallizable fragments.³¹

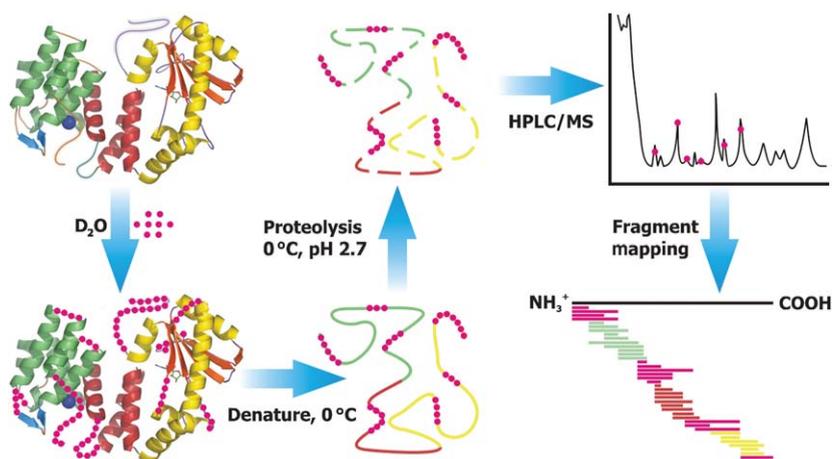


Figure 3 A schematic depiction of the hydrogen/deuterium-exchange MS (DXMS) procedure to aid the design of DNA constructs by identifying unstructured regions in the protein target. After establishment of initial protein fragmentation maps, the protein of interest is incubated in deuterium oxide (D_2O) for 10s at $0^\circ C$, leading to rapid exchange at solvent exposed amide nitrogens. After rapid denaturation in an acidic quench solution, deuterated samples are proteolyzed by brief exposure to immobilized pepsin. Proteolytic fragments are separated by HPLC, analyzed by MS, and disordered regions are localized by interpretation of amide hydrogen/deuterium exchange maps.^{30,31}

3.38.3.1.2 Affinity chromatography: finding the right tag

To aid in purification, recombinant protein targets are usually fused to an affinity tag, which can be added to the N- or C-terminus of the target sequence, or, if necessary, inserted within the sequence. To maximize the crystallization potential of the recombinant target, emphasis should be placed on diversification when selecting affinity tags. A vast number of affinity tags are available, and an ideal choice will depend on numerous factors including preferred tag location, tag size (ranging from a few residues to >100 kDa), and the desired balance of auxiliary tag characteristics such as gene expression and solubility of the tagged protein, ease of purification, overall purity of the eluted sample, or secreted expression.¹² Oligohistidine tags (His-tags) are often a first choice because of their relatively low cost, small size, and ease of use.³² The incorporation of a His-tag permits simple one-step purification using an immobilized metal-affinity chromatography (IMAC) resin such as Ni²⁺-nitrilotriacetate, but a second ‘polishing step’ using size-exclusion or ion-exchange chromatography is often needed to eliminate contaminating host proteins that have a natural affinity for IMAC resins. If subsequent tag removal is desired (e.g., if it is suspected that the tag might be interfering with crystallization or protein activity), vectors such as Qiagen’s pQE-30 Xa are available that allow incorporation of a cleavable tag containing a flanking protease recognition site. A recent comparative study of different affinity tags has touted the use of the *StrepII* tag (Trp-Ser-His-Pro-Gln-Phe-Glu-Lys), which binds a modified streptavidin-coupled matrix (*Strep-Tactin*).³³ According to this study, the *StrepII* tag affords much higher purity over the His tag at a comparable cost. Large affinity tags containing fusion proteins, such as the maltose-binding protein,³⁴ thioredoxin,³⁵ or glutathione-*S*-transferase³⁶ can improve folding and enhance solubility of the coupled target. Moreover, the growing number of chimeric target-fusion crystal structures that are being reported suggests that fusion proteins can be useful tools for crystallization.³⁷

3.38.3.1.3 Choosing a suitable expression system

Once the gene constructs have been designed, choosing the best expression system can be an equally troublesome procedure. The utilization of *Escherichia coli* as an expression host has long been the method of choice because it provides a fairly robust and easy means for producing recombinant proteins at a minimal expense. Despite these advantages, it should be pointed out that a large percentage (in some cases more than 50%) of cytosolic proteins encoded by bacterial and archaeal genomes appear to be insoluble when expressed in *E. coli*.³⁸ Production of eukaryotic proteins can be particularly problematic if posttranslational modifications such as phosphorylation or glycosylation are required for proper folding and activation of the recombinant target.³⁹ Eukaryotic expression systems offer the advantage of having the capability to phosphorylate and glycosylate recombinant proteins, but to various degrees.⁴⁰ Of these, the baculovirus system is particularly suitable to the high-throughput needs of an SBDD program, as insect cells are typically more durable than mammalian cells.¹³ Recently, the yeast *Pichia pastoris* has been genetically engineered into a ‘humanized’ form with a full complement of *N*-glycosylation machinery,⁴¹ a development which paves the way for large-scale production of recombinant human proteins with immediate applications ranging from protein therapeutics to SBDD.

To generate expression vector clones, newer cloning systems including the Novagen pTriEx, Invitrogen Echo, or Invitrogen Gateway kits are advantageous in that they allow the testing of multiple affinity tags in both *E. coli* and insect cells with minimal subcloning.⁴² The production of misfolded protein can often be alleviated by utilizing chaperones and foldases⁴³ or factorial screening methods.⁴⁴ Coexpression of the target with stabilizing binding partners is now possible with newer expression systems such as Novagen’s Duet coexpression vectors, which allow the simultaneous expression of up to eight proteins in the same cell. There are multiple cases illustrating the benefit of binding-partner coexpression if one protein component is insoluble or unstable when expressed on its own. For example, a complex of the polyomavirus internal protein VP2/VP3 with the pentameric major capsid protein VP1 was successfully prepared and crystallized only after coexpression of the components in *E. coli*. Coexpression was essential to obtain the complex since VP2 and VP3 fragments were found to be insoluble when expressed independently.⁴⁵

3.38.3.2 Milestone 2: From Protein to Crystal

The protein-to-crystal milestone can represent the most formidable challenge in SBDD. Biological macromolecules are complex and dynamic entities, with their own unique set of physicochemical properties, and no universal recipe exists for their crystallization. A purified sample may seem at first to be ‘intrinsically noncrystallizable,’ and a great deal of time and resources may be required to find an optimal set of crystallization conditions or to construct a suitable variant that produces high-quality crystals. Several strategies for accelerating this phase of the SBDD pipeline are outlined in the following subsections.

3.38.3.2.1 Solubility optimization: the first hurdle between purification and crystallization

Once a successful purification protocol has been established, it is frequently the case that the final elution buffer is inappropriate for initiating crystallization trials. Moreover, it is not unusual to observe that the solubility of the recombinant target in standard elution buffers is much lower than the protein concentration normally required for crystallization (usually 5–20 mg mL⁻¹). Even though Franz Hofmeister observed the differential solubility of proteins in various salts more than 125 years ago^{46,47} and a large amount of subsequent research has been conducted on this phenomenon,⁴⁸ protein solubility optimization can still be regarded somewhat as a ‘trial-and-error’ science because proteins are complex biological polyelectrolytes with fairly unpredictable solubility characteristics. The ordering of anions and cations according to their general ability to stabilize the structure of proteins has been referred to as the Hofmeister Series. Ammonium sulfate, which is a commonly used precipitant in protein crystallography, tends to both stabilize proteins in the folded state and drive them out of solution, while guanidinium chloride has the opposite tendencies. The differential effect of salts is one of the most important variables to screen when searching for conditions that enhance solubility of an overexpressed protein. Cosolvent additives, which can modulate protein solubility through direct interactions with the protein or by modifying the surrounding water structure,⁴⁹ are also worthwhile to explore for particularly recalcitrant proteins. Glycerol can enhance solubility without denaturing proteins.^{48,50} Detergents can improve solubility by binding to hydrophobic surface patches on proteins.⁵¹ Additional cosolvents such as small organic molecules, polyvalent ions, sugars, and polyhydric alcohols can also help to improve protein solubility and stability through different effects.^{52–56} Screening for an ideal solubility buffer composition is therefore a complex multiparameter puzzle. Lindwall and co-workers⁵⁷ were one of the first groups to approach this problem by developing a sparse-matrix solubility screen that could be used on crude cell extracts carrying the recombinant protein. The screen is empirically designed based on known solubility enhancers, and has produced favorable results in our laboratory for proteins that appeared otherwise insoluble in standard lysis and purification buffers. Alternative solubility screening methods employing dynamic light scattering or photometric analysis have been developed specifically for crystallography applications.^{58,59}

3.38.3.2.2 Protein characterization: fit for crystallization?

Protein purity is generally regarded as a critical determinant for successful crystallization. Of course there are always exceptions to this rule, as some proteins exhibit the propensity to crystallize even in the presence of high levels of impurities. The earliest account of successful protein crystallization by Friedrich Hünfeldt in 1840 reported on the growth of ‘blood crystals’ from earthworm (earthworm hemoglobin) upon slow dehydration of raw blood samples.⁶⁰ Indeed, many of the crystal specimens that supplied the seminal macromolecular x-ray studies of the early 1930s were coaxed from crude biological mixtures.⁶¹ As an example from our own laboratory, we have witnessed the remarkable propensity of one glycoprotein isozyme to crystallize out of an equal mixture of two different forms.⁶²

As a rule-of-thumb, however, protein purity should be regarded with utmost stringency when preparing samples for crystallization trials. Purified protein samples should not only be as free as possible from macromolecular contaminants (proteins, DNA/RNA, complex carbohydrates, etc.), but should also be chemically and conformationally homogeneous, exhibit a monodisperse size distribution, and be free of any denatured species or other microheterogeneities that may preclude or adversely affect crystallization. Automated assessment of purity can be carried out with high-throughput SDS-PAGE⁶³ and matrix-assisted laser desorption ionization (MALDI) mass spectrometry.^{64,65} Dynamic light scattering (DLS) measurements can be used to verify sample dispersity and this method has proven effective as a predictive tool for assessing the crystallizability of macromolecules.^{66–68} It was estimated in one study that as much as 70% of proteins exhibiting a monodisperse size distribution in a light scattering experiment will crystallize using a standard 48–96 condition sparse-matrix crystallization screen.⁶⁹

3.38.3.2.3 Dealing with heterogeneity

Posttranslational modifications (PTMs) are a major source of heterogeneity for proteins, and at least 150 different kinds of PTMs (e.g., glycosylation, methylation, phosphorylation, *S*-thiolation, etc.) are known,⁷⁰ with each affecting characteristics such as molecular weight, charge, and solubility of the modified protein. Although a PTM may be small, the modification may serve a physiological role and have a drastic impact on the overall conformation of the protein. From a crystallization standpoint, these different modified forms can be perceived as entirely foreign particles, and thus have similar adverse effects on the crystallization process. Chromatographic methods, enzymatic treatment, or mutagenesis can help to reduce or eliminate heterogeneity.

3.38.3.2.3.1 Glycosylation and phosphorylation

Glycosylation can be a particularly problematic source of heterogeneity when eukaryotic expression systems are used. The negative influence of attached oligosaccharides on the crystallizability of glycoproteins presumably stems from elevated surface entropy due to conformational flexibility of the oligosaccharide chains, as well as differences in glycan chain length, branching, and sugar composition. Removal of the attached oligosaccharides by enzymatic deglycosylation has been suggested as a way to overcome this problem,^{71,72} and indeed several reports have emphasized the importance of complete deglycosylation for obtaining suitably diffracting crystals.^{73–75} For complete removal of asparagine-linked (N-linked) oligosaccharides, peptide-*N*-glycosidases such as PNGase F or PNGase A are often used, which also cause deamidation of the asparagine residue to aspartic acid.⁷⁶ In our experience, however, leaving the innermost sugar residue attached to the protein may offer additional possibilities for sugar-mediated crystal-contact formation,⁶² and different deglycosylation strategies (selective versus complete deglycosylation) should therefore be devised (Figure 4). As an example for selective *N*-deglycosylation, partial digestion might be attempted with selective endoglycosidases (such as endo F₁–F₃ or endo H), which leave the innermost *N*-acetyl glucosamine (GlcNAc) of the *N*-linked diacetylchitobiose glycan core attached to the modified Asn. In cases where the GlcNAc is fucosylated, a combined treatment with α -fucosidase might be warranted. Unfortunately, there is no enzyme comparable to PNGase F or PNGase A for removing *O*-linked glycans from Ser/Thr residues. Monosaccharides must be sequentially trimmed off by a series of exoglycosidases until the Gal-(1,3)-GalNAc core remains, at which point *O*-glycosidase can be used to remove the core structure without chemically modifying the Ser/Thr residue.⁷⁷ Modifications to the core structure can often block the action of *O*-glycosidase and may require the use of additional glycosidases such as α -2-(3,6,8,9)-neuraminidase.⁷⁸ It should be noted that a nonspecific galactosidase or β -(1,3)-galactosidase can be used to hydrolyze the galactose from the core glycan, leaving the *O*-linked GalNAc attached to the Ser/Thr and thereby providing an alternative to complete sugar removal.

Mutagenesis can be a suitable alternative to enzymatic or chromatographic approaches for generating homogeneous samples for crystallographic studies. For example, substitution of Asn or Ser/Thr in Asn–Xaa–Ser/Thr sequons can prevent *N*-glycosylation. Heterogeneous phosphorylation can be a problem, for instance, when expressing kinases in insect cells, and enzymatic treatment with alkaline phosphatase or phage λ protein phosphatase has been used successfully for obtaining completely dephosphorylated protein for crystallographic studies.^{79–81} Alternatively, mutation of Ser/Thr- or Tyr-phosphorylation sites will eliminate heterogeneous phosphorylation altogether. When phosphorylation plays a critical modulatory role in the target protein's activity, a phosphorylated conformation might be of interest for SBDD. In such cases, generating a 'phosphorylated' conformation can often be accomplished by replacing the phosphorylated Ser/Thr residue with Asp or Glu, with the electronegative side chain carboxylate moiety acting as a phosphate mimic.^{80,81}

3.38.3.2.3.2 Isoelectric heterogeneity

Microheterogeneities such as isoelectric heterogeneity can be overcome by additional purification steps such as ion-exchange chromatography or preparative isoelectric focusing techniques. The potential impact of microheterogeneity on protein crystallizability is exemplified by the work of Prongay and colleagues,⁸² who crystallized a complex of human immunodeficiency virus capsid protein p24 with its antigen-binding fragment (Fab). In this work, recombinant p24 was purified to homogeneity and chromatographically separated into distinct isoelectric species, only one of which was able to crystallize in complex with the Fab. Moreover, different crystal forms of the p24:Fab complex could be obtained when different isoelectric species of the Fab were used in combination with the crystallizable p24 fraction.

3.38.3.2.3.3 Utilizing combined approaches to eliminating heterogeneity

A recent success story employing several of the methods described in this and preceding sections is the expression and structure determination of a truncated and deglycosylated form of human ACE from testes (tACE).^{10,11,83} A truncated form of tACE, lacking the N-terminal 36 residues and the C-terminal transmembrane domain was expressed in the presence of *N*-butyl-deoxynojirimycin, an α -glucosidase I inhibitor. After expression, the mutant tACE (tACE Δ 36NJ) was treated with endoglycosidase H to selectively remove all but the innermost GlcNAc residue at each *N*-glycosylation site. Although the expression was low, tACE Δ 36NJ was shown to be homogeneous on SDS-PAGE, fully active in enzymatic assays, and produced crystals diffracting to 2.0 Å resolution. This led to the long-awaited structural elucidation of a human ACE in complex with captopril and related antihypertensives (as described previously in this chapter), and paved the way for design of next-generation inhibitors (Figure 1). The glycosylation sites (Asn–Xaa–Ser/Thr) on tACE Δ 36 were mutated systematically by replacing Asn with Gln.⁸⁴ Two of the Asn \rightarrow Gln tACE Δ 36 mutants exhibited

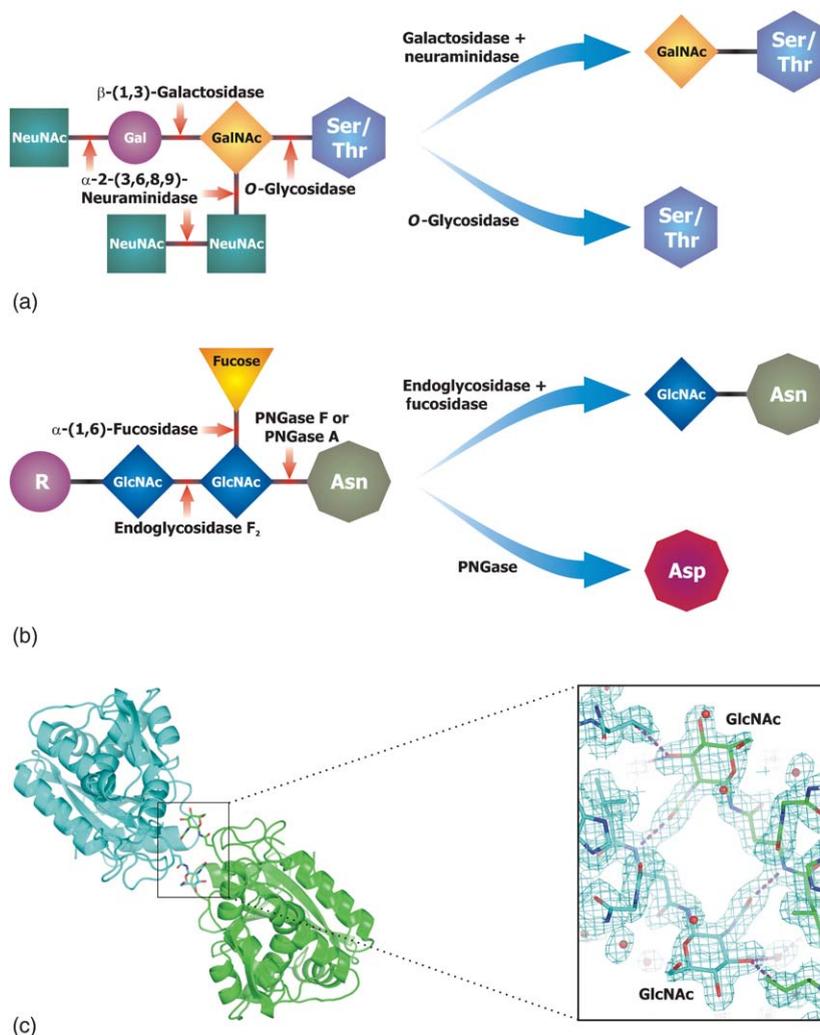


Figure 4 Examples of alternative deglycosylation strategies for *O*- and *N*-linked glycans. Cleavage sites of various glycosidases are indicated with red arrows. Sugar abbreviations used: NeuNAc, *N*-acetylneuraminic acid; Gal, galactose; GalNAc, *N*-acetylgalactosamine; GlcNAc, *N*-acetylglucosamine. (a) A trisialated *O*-linked glycan is digested with a combination of neuraminidase and galactosidase to yield a glycoprotein variant containing a residual *O*-linked GalNAc. A second protein variant is produced by complete glycan removal with *O*-glycosidase. (b) An *N*-linked biantennary or high-mannose glycan is trimmed with endoglycosidase and fucosidase to generate a glycoprotein with one remaining GlnNAc residue attached. Alternatively, treatment with peptide-*N*-glycosidase (PNGase) facilitates complete glycan removal with concomitant deamidation of the Asn to produce a variant bearing an Asp residue. (c) The crystal structure of pokeweed antiviral protein (PAP) reveals the potential benefit of *O*- and *N*-linked monosaccharides in promoting crystallization of glycoproteins.⁶² Depicted are two crystallographically related PAP monomers engaged in a crystal contact via *N*-linked GlcNAc monosaccharides. The isoform lacking this modification failed to crystallize.

higher expression levels and produced crystals isomorphous with the tACE Δ 36NJ construct, although the resolution was not as good (~ 3.0 Å).

Another relevant success story involves the rational re-engineering of the anticancer target, human urokinase, for crystallographic studies.⁸⁵ Preliminary crystals of urokinase in complex with an inhibitor diffracted to 2.5 Å resolution, yet when the structure was solved, it was discovered that the active site was effectively shielded by tight intermolecular packing in the crystal. This unfortunate packing arrangement precluded the diffusion of small molecule inhibitors into the urokinase active site by crystal soaking methods. Using the initial crystal structure as a guide, all disordered or highly flexible polypeptide regions (as judged by missing electron density or high atomic temperature

factors) were deleted from the gene construct. Moreover, a free Cys residue, which was exposed by the truncation, was mutated to Ala, and an additional Asn→Gln point mutation was introduced to remove an *N*-glycosylation site. The resulting modified urokinase, termed micro-urokinase, was subsequently crystallized in a better-diffracting crystal form, which was suitable for high-throughput soaking experiments and rational drug design.

3.38.3.2.4 The protein as a variable

The preceding section dealt with different strategies for improving crystallization probability by overcoming heterogeneity. Unfortunately, reality can (and often will) painfully dictate that some proteins, even in their purest, most monodisperse and homogeneous form(s), simply won't crystallize. For these recalcitrant types, fruitless screening of thousands upon thousands of crystallization conditions covering wide expanses of crystallization space will only galvanize the truth. The results of many structural genomics initiatives around the world have confirmed that the majority of proteins that have the propensity to crystallize will do so over an astonishingly narrow range of crystallization conditions.^{86–88} Consequently, there is a growing emphasis on screening the protein target as a variable in protein crystallization, for example, either by homolog/ortholog screening, site-directed mutagenesis, directed evolution, or by chemical means.

The overall probability of crystallization of a target is exponentially related to the number of protein variants utilized. Assuming complete independence of variants, the overall crystallization probability, p_T would be $1 - (1 - p_{ave})^n$, where p_{ave} is the average probability of crystallizing one of the n variants. As an example, if each variant of a target only had a 20% chance of crystallizing, but 20 variants were constructed, the overall probability of successful crystallization would be $1 - (1 - 0.20)^{20}$,²⁰ or nearly 99%. Wayne Hendrickson and colleagues were the first to systematically analyze the crystallization probability enhancement of different protein modifications and multiple combinations thereof on a specific target, gp120 (the envelope glycoprotein of type 1 human immunodeficiency virus), which had previously resisted crystallization.⁸⁹ By exploring different deglycosylation protocols, substituting different surface loops with tripeptide linkers of Gly-Ala-Gly, producing N- and C-terminal deletion mutants, and reducing conformational heterogeneity by using different protein ligands such as its receptor CD4 and Fabs from conformationally sensitive monoclonal antibodies, they were successful in growing six different types of gp120 crystals from 18 different variants. One of these, the ternary complex between gp120, CD4, and the Fab of the human neutralizing monoclonal antibody 17b, was determined at 2.5 Å resolution.⁹⁰

3.38.3.2.4.1 Homolog screening and DNA shuffling

The value of using homologs for crystallization screening purposes was pioneered decades ago by John Kendrew, whose initially unsuccessful attempts at crystallizing horse heart myoglobin were rescued by choosing sperm whale as an alternative source for material.^{91,92} In another piece of seminal work, Herman Watson and colleagues used protein from different sources to crystallize enzymes from the glycolytic pathway.⁹³ Systematic variation in the species of origin was also instrumental in the crystallization of the transcription initiation TATA-binding protein.⁹⁴ Recent structural proteomics data have corroborated this notion by showing that inclusion of homologous proteins of a target drastically increases crystallization success rates.⁹⁵ In a classic set of papers demonstrating the effectiveness of iterative structure-based drug discovery using a target homolog, the Agouron group designed potent antitumor agents targeting human thymidylate synthase by using the crystallizable *E. coli* enzyme as a surrogate.^{96–98} An extension of this approach is exemplified by the SBDD of selective cyclin-dependent kinase (CDK) inhibitors. Because efforts to obtain crystals of CDK-4 for design of CDK-4-selective inhibitors were unsuccessful, Merck scientists mutated the ATP-binding pocket of the crystallizable CDK-2 homolog to create a CDK-4 active-site mimic. This approach resulted in the successful design of a CDK-4-specific inhibitor.⁹⁹ During the recent severe acute respiratory syndrome (SARS) outbreak, our group was able to integrate the structural data which we obtained on the main proteinases of human coronavirus 229E and transmissible gastroenteritis (corona)virus, with molecular modeling and biochemical analyses, to propose that derivatives of AG7088 (an inhibitor of the distantly related human rhinovirus 3C proteinase) would be promising lead candidates for the design of anti-SARS drugs targeting the SARS coronavirus main proteinase.¹⁰⁰

When crystallization screening with homologs fails to generate hits, the directed evolution approach of 'DNA shuffling' may provide the necessary breakthrough (Figure 5). The DNA shuffling method relies on homologous recombination during the PCR reassembly of gene fragments from multiple parents to generate crossovers at points of high sequence identity. The procedure can generate a vast library of chimeras, which can be rapidly screened in parallel for expression, solubility, and activity. Sequence variants that meet or exceed solubility and activity criteria can then undergo crystallization trials. This strategy has recently paid off for Keenan and colleagues who implemented a directed evolution program for engineering a glyphosate *N*-acetyltransferase (GAT) superenzyme to confer glyphosate resistance

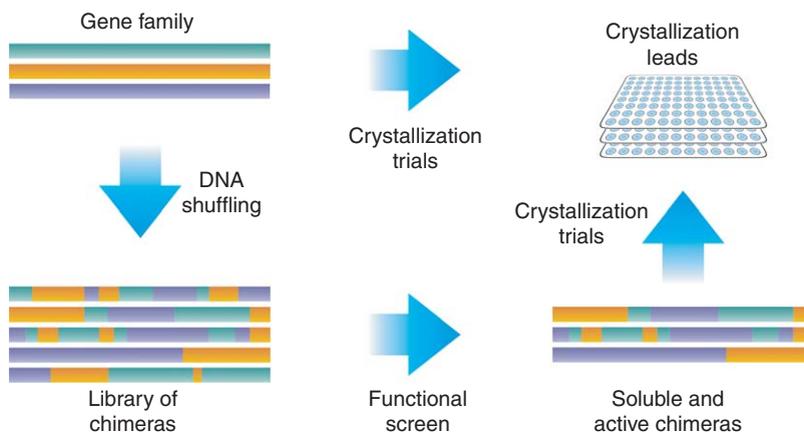


Figure 5 A diversified crystallization strategy employing a combination of homolog screening with DNA shuffling to improve success probability.

in transgenic plants.¹⁰¹ Initial attempts to crystallize three wild-type GAT enzymes, both in the presence and absence of ligands, failed to produce crystals suitable for structure determination. To overcome this bottleneck, multiple rounds of DNA shuffling were carried out between three wild-type GAT genes in conjunction with high-throughput screening to identify GAT chimeras with decent expression, strong enzymatic activity, and high solubility. Eight randomly chosen GAT variants that met the above criteria were selected from the library, with each containing 5–25% exchanged amino acid content relative to the wild-type enzymes. Utilizing the same crystallization procedure as for the wild-type enzymes (96-condition sparse matrix screen conducted both in the presence and absence of ligands), two of the eight GAT chimeras produced suitable crystals, and the crystal structure of one of them was solved. Not surprisingly, the resulting 1.6 Å crystal structure showed that some of the exchanged regions in the shuffled GAT variant were involved in crystal contact formation.¹⁰¹

3.38.3.2.4.2 Metal-mediated crystallization

While orthologs and DNA-shuffled chimeras may exhibit multiple regions of sequence variability on their protein surfaces, which could provide unique areas for crystal packing, a single point mutation is often sufficient to drastically enhance the crystallizability of a target. Perhaps the earliest example demonstrating the effectiveness of this method was the rational genetic engineering of crystal contacts in human H ferritin.¹⁰² The Czech physiologist Vilem Laufberger noted in 1937 that ferritin could be crystallized in situ by adding droplets of concentrated cadmium sulfate to fresh slices of horse spleen.¹⁰³ The crystal structures of horse spleen ferritin, rat liver ferritin, and recombinant rat L-chain ferritin – all crystallized in the presence of CdSO_4 – revealed intermolecular crystal packing via double Cd^{2+} bridges, with pairs of aspartate and glutamine side chains from neighboring ferritin molecules forming ligands to the metal ions.^{104–106} Human and other H-rich ferritins, on the other hand, immediately form amorphous precipitate upon the addition of Cd^{2+} , even at low concentrations.¹⁰⁷ Sequence comparisons revealed that human H ferritin contained a lysine in place of the Cd^{2+} -coordinating glutamine residue. By constructing a Lys → Gln mutant of recombinant human ferritin, large well-diffracting crystals isomorphous to those of the horse and rat ferritins were obtained. Surprisingly, these crystals could only be grown in the presence of Ca^{2+} , although subsequent structure determination revealed that the two Ca^{2+} ions formed an identical intermolecular coordination arrangement as Cd^{2+} in the horse and rat ferritins.¹⁰²

The general utility of metal ions and organometallic compounds to induce macromolecular crystal growth by forming metal-mediated crystal contacts cannot be overstated, and a growing number of success stories documenting this fact continue to fill the literature. The practice of using metal ions in protein crystallization dates as far back as 1925 when John J. Abel – perhaps fortuitously at first – used Zn^{2+} to obtain rhombohedral crystals of insulin.¹⁰⁸ In our laboratory, Zn^{2+} remains the divalent cation of choice for crystallization additive screening, with examples including the Zn^{2+} -mediated crystallization of the macrophage infectivity potentiator protein from *Legionella pneumophila*,¹⁰⁹ and a human carbonic anhydrase (TH, unpublished) (Figure 6). It should be emphasized, however, that Zn^{2+} will not be the best crystallization additive for every protein, and provided that enough protein sample is available, parallel screening of different metal salts is recommended. To illustrate this, a study examining the influence of various divalent cations on the crystallization of isoleucine/valine-binding protein and leucine-specific binding protein led to

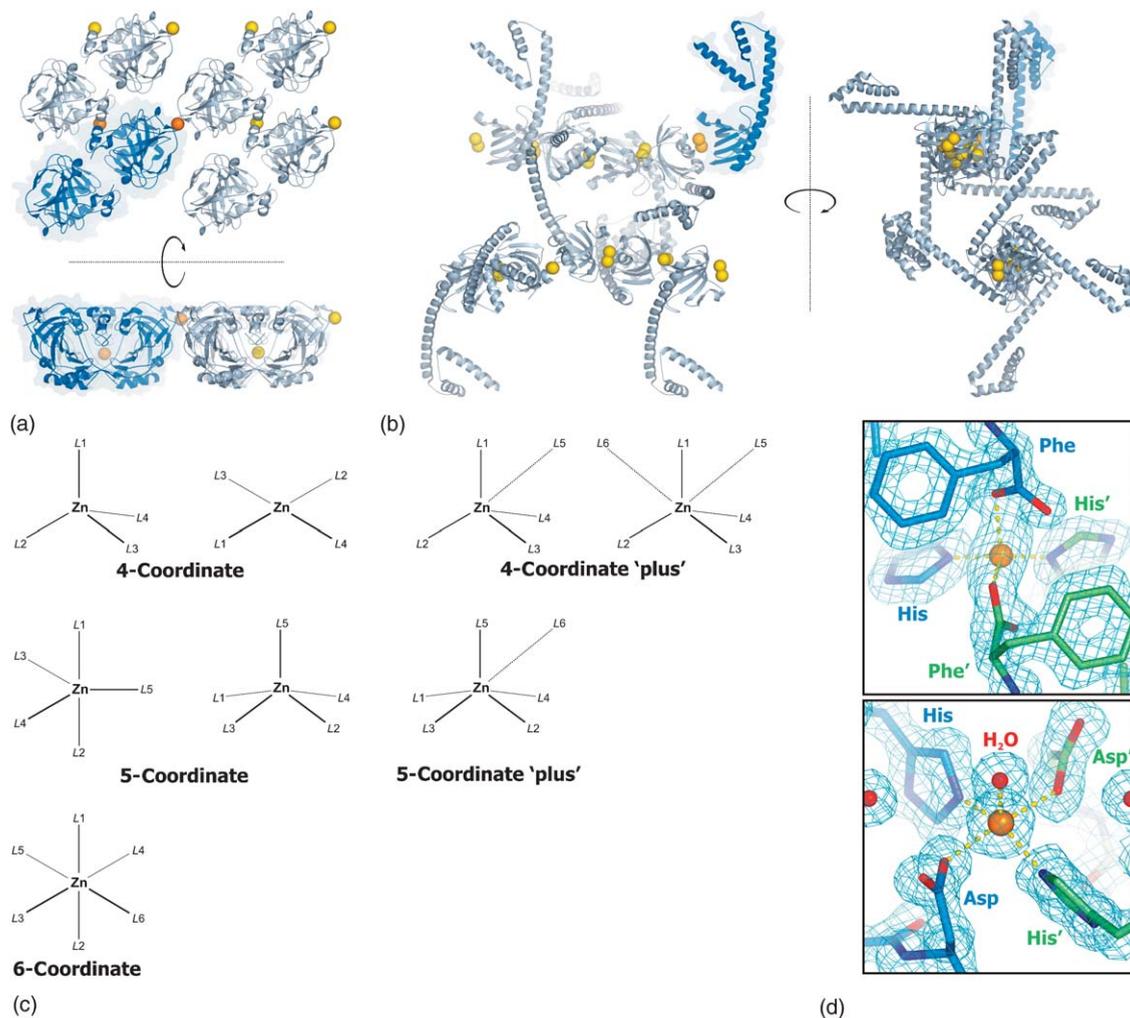


Figure 6 Zn²⁺-mediated crystallization of drug targets, exemplifying the general utility of metals in crystallization additive screening. (a) Orthogonal views of the crystal lattice in P1 crystals of human carbonic anhydrase (TH, unpublished data). The crystallographic asymmetric unit contains two monomers (highlighted in blue), and the two Zn²⁺ ions involved in crystal contact formation are depicted as orange spheres. The tight crystal packing promoted by the intermolecular metal-bridging is visibly apparent. (b) Orthogonal views of the crystal lattice in tetragonal crystals of the macrophage infectivity potentiator protein (Mip) from *Legionella pneumophila*.¹⁰⁹ In this case, a dinuclear Zn²⁺ crystal-contact bridges the catalytic domains of symmetry-related monomers in an otherwise loosely packed crystal environment. In both of the above cases, the presence of Zn²⁺ in the crystallization buffer was mandatory for crystallization. (c) The various coordination states and ligation geometries observed for Zn²⁺ in protein crystals. 4-Coordinate and 5-coordinate states can be additionally decorated with weaker long-range interactions (denoted by 'plus'). (d) Two examples of Zn²⁺-mediated crystal contacts observed at high resolution (TH, unpublished data). In the top example, the Zn²⁺ establishes a 4-coordinate tetrahedral crystal contact involving the C-terminal carboxylate moieties (Phe) and His side chains of symmetry-related monomers. The bottom example illustrates a 5-coordinate square-pyramidal Zn²⁺ crystal-contact ligated by the identical His/Asp residues in adjacent monomers. A water molecule occupies the apical position.

the observation that different metals promoted different crystal forms, with Cd²⁺ producing the best results.¹¹⁰ This and a subsequent study with histidine-binding protein from *Salmonella typhimurium* also found that the optimal metal concentration for crystallization fell within a very narrow range.¹¹¹ It is as yet impossible to predict a priori which type of metal or organometallic complex, nor which concentration range (μM – M), will be most suitable for crystallizing a given protein; however, the remarkable propensity of Zn²⁺ to exhibit a wide array of coordination geometries (4-, 5-, and 6-coordinate geometry both with and without additional weak interactions) and a broad range of potential protein ligand functions (Asp, Glu, His, Cys, main-chain carbonyl, and C-terminal carboxylate),^{112,113} together with its utility

for multiple-wavelength anomalous dispersion phasing methods,¹¹⁴ make it a logical first choice when metal-additive crystallization screening is considered (Figure 6). The Pfizer group has recently proposed engineering carboxylic protein surface residues (Asp and Glu) into proteins as a crystallization strategy to enable Zn^{2+} -mediated crystal contacts.¹¹⁵

3.38.3.2.4.3 Mutational strategies I: rational surface engineering

Since the original work by Lawson and co-workers on human H-chain ferritin,¹⁰² a rapidly growing number of examples are appearing in which mutated proteins have crystallized when the wild-type counterparts have not.^{116,117} In a classic example of ‘serendipity dictating crystallizability,’ the bacterial chaperonin GroEL was crystallized only after two mutations were accidentally incorporated by PCR.^{118,119} In an elegant study by Ernest Villafranca’s group at Agouron, nonconserved surface residues of human thymidylate synthase were rationally mutated to residues with altered polarity or charge.¹²⁰ All resulting point mutants had altered crystallization behavior, were generally more crystallizable than the wild-type protein, crystallized under unique conditions, and produced novel crystal forms that were of higher quality and more useful for inhibitor soaking studies. An analogous study by the Roche group with point mutants of the *E. coli* DNA gyrase B subunit produced similar improvements in crystallizability and crystal quality.¹²¹

The mind-boggling number of possibilities that present themselves when considering site-directed mutagenesis as a tool for crystallizing a new target raises a few fundamental questions: which sites in the protein sequence should be chosen for mutagenesis, and what kind of mutation will produce the best results? For the most part, the strategy has until now remained trial-and-error in nature, and the relative paucity of empirical data on the subject still leaves much to be answered. For example, solubility and crystallization screening of about 30 different hydrophobic \rightarrow Lys point mutants of the catalytic domain of HIV integrase led to crystallization of the Phe185Lys mutant.¹²² In another case, obtaining of diffraction-quality crystals of the mycobacterial outer membrane channel was possible only after preparing about a dozen point mutants – either within zones of interspecies sequence variation or at predicted surface regions – of which an Ala96Arg mutant produced crystals diffracting to 2.5 Å resolution.¹²³ A Trp100Glu point mutation in wild-type human leptin produced a soluble protein that crystallized and allowed structure determination to 2.4 Å.¹²⁴ A Lys133Ile mutant of cyclophilin D gave crystals that diffracted to 1.7 Å resolution.¹²⁵ What becomes clear at this point is there has been no real consensus on what works best; however, recent work by Zygmunt Derewenda’s group has illustrated the consistent effectiveness of incorporating mutations that reduce conformational surface entropy.^{117,126} The underlying concept stems from the idea that surface residues with high conformational entropy create an ‘entropic shield’ that prevents the formation of stable intermolecular interactions required for crystal contact formation. Therefore, the selective replacement of conformationally labile surface residues with small amino acids, such as Ala, could reduce the entropic penalty of crystal contact formation in these regions. Lysine and glutamate residues have been identified as good candidates for mutagenic replacement because they are usually located on the protein surface,¹²⁷ their side chains are typically characterized by high conformational entropy,¹²⁸ and they are disfavored at protein–protein interfaces.¹²⁹ The hypothesis was emphatically validated by mutational work on RhoGDI, the human Rho-specific GDP dissociation inhibitor, of which several mutants were constructed (each containing between one and four Lys \rightarrow Ala and/or Glu \rightarrow Ala exchanges) and subjected to crystallization trials.^{130,131} Almost all mutants exhibited altered crystallization properties and several new crystal forms were obtained, with one Glu154Ala/Glu155Ala double-mutant producing crystals that diffracted to 1.25 Å resolution.¹³¹ Crystal structure determinations of the different mutants revealed that the mutated epitopes participated directly in crystal contact formation. This surface engineering strategy has since produced successful results for a number of proteins, and the current trend involves making multiple mutations within short Lys/Glu-rich clusters.^{117,126} A recent example of this approach toward SBDD is provided by work from Merck scientists who investigated the tyrosine kinase domain of the insulin-like growth factor-1 receptor (IGF-1R), an anticancer target. Crystals of the unphosphorylated apo form of the IGF-1R kinase domain diffracted to only 2.7 Å resolution, a resolution which is too low for high-throughput structure-based lead optimization. To overcome this hurdle, three different multiple mutants were prepared, Lys1025Ala/Lys1026Ala, Glu1067Ala/Glu1069Ala, and Lys1237Ala/Glu1238Ala/Glu1239Ala, with the second double-mutant producing high-quality crystals diffracting to 1.5 Å resolution.¹³² In the event that the solubility of the mutant protein is critically compromised, alternative mutations to polar residues with less conformational entropy and/or smaller size (e.g., Lys \rightarrow Arg; Lys \rightarrow Asp) can also help to promote crystallization.^{133,134} The concept of rationally designing proteins to crystallize has been extended by Wingren and colleagues who have shown it is possible to replace short stretches of residues in β -strand-containing proteins with so-called packing ‘cassettes’ – crystal packing motifs that promote crystallization by generating specific crystal packing interactions.¹³⁵ As with any mutational work, the functional integrity of the mutant proteins needs to be evaluated prior to SBDD.

3.38.3.2.4.4 Mutational strategies II: directed evolution

Directed evolution approaches to producing soluble and crystallizable targets have recently emerged as powerful alternatives to site-directed mutagenesis.¹³⁶ These methods allow large mutant libraries to be generated through error-prone PCR and DNA shuffling. Mutants can be expressed as green fluorescent protein (GFP) fusions and visibly assayed for solubility and proper folding. The great advantage of the GFP-based directed evolution approach compared to rational surface-engineering is that it eliminates all of the guesswork related to mutant design and samples mutational space in a much more efficient way. Yang and co-workers demonstrated the superior utility of GFP-based directed evolution for obtaining soluble and crystallizable mutants of the protein RV2002 from *Mycobacterium tuberculosis*.¹³⁷ Mutants of RV2002 generated by GFP-based directed evolution had solubilities of at least $\sim 15 \text{ mg mL}^{-1}$, compared to the wild-type protein, which was expressed in *E. coli* only as insoluble inclusion bodies. An important observation raised by the authors in their crystallographic study was that the underlying contributions of the individual mutations toward the enhanced solubility of the Ile6Thr/Val47Met/Thr69Lys triple mutant appeared to involve a combination of altered intrinsic solubility and folding kinetics.¹³⁷ It is very likely that it would have been much more difficult, if not impossible, to rationally engineer soluble mutants as quickly or effectively.

3.38.3.2.4.5 Chemical modifications

An alternative strategy that deserves consideration before embarking on the more time-consuming approaches of ortholog screening, rational surface engineering, or directed evolution (or perhaps even as a final recourse in the event that the above methods fail) is to chemically modify the target protein prior to crystallization. Reductive alkylation of the protein sample with formaldehyde and dimethylamine-borane complex (DMAB) has proven to be the most successful chemical treatment strategy established thus far for crystallization of recalcitrant samples. Although the technique was originally used to improve the quality of poorly diffracting crystals, its use as a tool for de novo crystallization was pioneered by Ivan Rayment for the structure determination of myosin subfragment 1,¹³⁸ and a methylation protocol of general utility has been subsequently published.¹³⁹ The net consequence of this treatment is the dimethylation of all solvent-accessible lysine side chains as well as the N-terminal amino group. The resulting charge of the modified protein is not changed, however, its isoelectric point may be shifted slightly. As with surface-entropy-reducing mutagenesis, it is believed that the crystallizability of the chemically modified target may be enhanced as a result of reduced side chain entropy of dimethylated lysines. Additional studies have demonstrated that methylated proteins can produce novel crystal forms differing from unmodified samples,^{140,141} and that selenomethionine-substituted specimens as well as multimeric complexes are amenable to this form of chemical treatment.¹⁴² Other chemical approaches for the purpose of crystallization remain less explored, although a recent study has demonstrated the utility of deliberate oxidation in this regard. Working on a bacterial Ppx/GppA phosphatase homolog, Kristensen and co-workers were able to obtain two unique crystal forms after incubating their protein samples with 0.1% H_2O_2 for 1 h at room temperature prior to crystallization. The oxidized crystal forms were unique and were found to exhibit increased physical stability and better diffraction compared to the crystal form obtained with untreated samples.¹⁴³

3.38.3.3 Milestone 3: From Crystal to Structure

3.38.3.3.1 Postcrystallization treatments to improve diffraction quality

One of the major disappointments regularly encountered in an SBDD program employing crystallography is the realization that crystals of the protein target are of poor quality and unsuitable for diffraction studies. The phenomena of low-resolution or poor-quality diffraction (e.g., characterized by severe mosaicity or anisotropy) usually stems from loose molecular packing within the crystal and/or a high internal solvent content. These symptoms often arise from crystals that otherwise appear to be physically perfect when viewed under the light microscope, adding to a researcher's frustration and invariably leading to serious doubts about proceeding with the project. Before capitulating, however, there are several postcrystallization treatments at hand that can improve the diffraction quality of protein crystals, sometimes drastically.¹⁴⁴ The most successful methods currently employed for improving diffraction quality at this stage are crystal dehydration and 'crystal annealing.' The latter method arose as a way to repair crystal lattice damage inflicted by flash-cooling techniques (rapid cooling of crystals to cryogenic temperatures),^{145,146} which are routinely used to limit crystal radiation damage induced by exposure to high-intensity x-ray sources.^{147–155} Crystal annealing involves warming a flash-cooled crystal to room temperature followed by repeated flash cooling, and has been shown to improve crystal mosaicity and diffraction resolution.^{156,157} The method of crystal dehydration effectively reduces crystal solvent content, enforcing tighter crystal packing and frequently promoting remarkable improvements in diffraction quality. Several dehydration methods have emerged, ranging from slow controlled dehydration of the crystallization droplet over a period of days or weeks, to quick crystal soaks in a dehydrating solution for a few minutes.

A novel device has been described that allows accurate control of crystal water content by regulating the relative humidity of a gas stream enveloping the crystal.¹⁵⁸ This device, which is now available commercially, allows the diffraction properties of the crystal to be monitored in real time while the crystal is dehydrated until diffraction is improved. A recent study has demonstrated that the method of fast crystal dehydration coupled with crystal annealing can lead to astonishing improvements in diffraction quality, sometimes extending the diffraction limit of some apparently worthless crystal specimens by nearly tenfold.¹⁵⁹ Other postcrystallization treatments for improving diffraction resolution, including crystal soaking and crystal cross-linking, have been reviewed recently.¹⁴⁴

3.38.3.3.2 Cryocrystallography in practice

As mentioned, cryocrystallography is the method of choice when collecting diffraction data at high-intensity X-ray sources due to its dampening effect on the diffusion of harmful free radicals through the crystal, allowing most crystals to survive long enough in the x-ray beam for collection of a complete single-crystal high-resolution data set. A critical objective of cryocrystallography is the formation of amorphous ice within and around the crystal upon cryocooling, as crystalline ice formation yields spurious diffraction that obscures the useful protein diffraction. Amorphous ice formation can be facilitated by the addition of chemical cryoprotectants, such as glycerol or polyethylene glycol, to the crystallization mother liquor.^{154,160} Sometimes the crystallization mother liquor will already contain a cryoprotective formulation of reagents, thereby facilitating the cryocooling of crystals directly from their growth solutions. Several commercial crystallization screening kits containing cryo-ready reagents have been developed based on this strategy.¹⁶¹ One possible unwanted side effect with regard to SBDD, however, is the proclivity of cryoprotectant molecules to appear in the active sites of enzymes, in which case the screening of alternative cryoprotectants might be needed. A plausible workaround, which can eliminate the need for conventional cryoprotectants, is the use of cryoprotective oils such as Paratone-N or highly liquid paraffin oil, which work by forming a protective shield around the crystal.^{148,162,163} Another cooling method that circumvents the need for penetrative cryoprotectants is high-pressure crystal cooling, which was first explored as early as 1973¹⁶⁴ and recently revisited.^{165,166} A prototype pressure device has been constructed at the Cornell High Energy Synchrotron Source (CHESS) that allows for cryo-loop crystal mounting and high-pressure (200 MPa) cryocooling. Preliminary results with the device, which have shown significant improvement of diffraction quality for all protein crystals studied, are very encouraging.¹⁶⁶

3.38.3.3.3 Phasing methods for drug discovery

Once protein crystals can be grown reproducibly in a high-quality crystal form suitable for SBDD, the next major bottleneck is finding the solution to the phase problem. If coordinates of a related homolog or domain are available, model-based phasing using the molecular replacement method is the preferred strategy because of its relative speed.¹⁶⁷ Modern molecular replacement software implementing improved search features such as likelihood-enhanced rotation and translation functions,^{168,169} improved procedures for constructing molecular replacement search models,¹⁷⁰ and a burgeoning PDB databank of 3D macromolecular structure data,^{171,172} are literally pushing the boundaries of what can be accomplished with the method.

When preexisting structural data is absent or insufficient for molecular replacement phasing, other methods must be employed. Historically, the most commonly used structure-solving technique was multiple isomorphous replacement (MIR), which relies on the preparation of derivative crystals in which one or more types of heavy atoms are bound specifically and uniformly to the macromolecules within the crystal. The outcome of the MIR method is additionally contingent on the heavy-atom derivatives being truly isomorphous, i.e., bearing no (or at most, extremely limited) alterations in molecular structure and unit-cell dimensions. Successful phase determination by MIR is consequently rate-limited by the speed with which suitable heavy-atom derivatives can be obtained. The advancement of multiple-wavelength anomalous dispersion phasing^{114,173} and the construction of tunable synchrotron x-ray beamlines worldwide have made multiple-wavelength anomalous dispersion the method of choice for experimental phase determination, particularly for high-throughput (HT) SBDD applications. One of the big advantages multiple-wavelength anomalous dispersion phasing offers over MIR and related methods such as SIRAS and MIRAS (single or multiple-isomorphous replacement with anomalous scattering) is the reduction of systematic error since all data are measured on a single sample. The resulting phase angles are therefore more accurately determined and the resulting electron density maps are typically of higher quality than maps calculated with phases derived from isomorphous methods. Multiple-wavelength anomalous dispersion phasing requires the presence of a suitable number of atoms having an x-ray absorption edge in the energy range easily accessible by tunable synchrotron sources (typically $\lambda = 0.8\text{--}1.3\text{ \AA}$). Most multiple-wavelength anomalous dispersion phasing experiments are carried out using selenomethionine-substituted (SeMet) proteins,^{174,175} although sometimes SeMet multiple-wavelength anomalous dispersion phasing can be an

inappropriate method for phase determination.¹⁷⁶ In such situations, multiple-wavelength anomalous dispersion phasing can be carried out with bound elements of atomic numbers ~ 20 – 40 (or above ~ 60), as they have synchrotron-accessible absorption edges. These elements can often be found as natural metal cofactors in proteins (such as transition metals) or can be diffused into protein crystals using conventional heavy-atom soaking methods.^{177–180} Prescreening for the degree of SeMet or heavy atom incorporation into the target protein can be rapidly carried out by mass spectrometry.¹⁸¹ Alternatively, native polyacrylamide gel electrophoresis experiments allow for a quick and simple assessment of heavy atom binding, optimal binding concentrations, and impact on protein stability.¹⁸²

Explorations of novel approaches for introducing anomalous scattering elements into protein crystals for the purpose of multiple-wavelength anomalous dispersion phasing has led to a robust collection of new phasing alternatives. For example, chemically modified ligands can be used as rational phasing tools, such as halogenated nucleotides for multiple-wavelength anomalous dispersion phasing of nucleotide-binding proteins¹⁸³ or selenium-substituted saccharides for carbohydrate-binding proteins.¹⁸⁴ Multiple-wavelength anomalous dispersion phasing with krypton has been demonstrated as a feasible technique,¹⁸⁵ as has cryosoaking with halides^{186,187} or mono- and polyvalent cations.^{188,189} Several other unique procedures for producing derivatives have appeared over the years and continued exploration in this field will surely generate additional phasing alternatives. Moreover, modern developments in the area of utilizing soft x-rays ($\lambda = 1.5$ – 3.0 Å) for sulfur-based anomalous phasing – a technique pioneered more than 20 years ago by Hendrickson & Teeter for the structure determination of crambin¹⁹⁰ – may eventually circumvent the need for derivatization in most cases altogether.^{191–194}

3.38.3.4 Milestone 4: Structure to Lead

Once the crystal structure of a particular target has been solved, and before SBDD can begin, the crystal form should be thoroughly scrutinized for the presence of characteristic pathologies that might prolong or even prohibit the design process. Is the diffraction resolution high enough (i.e., a resolution of < 2.5 Å, or preferably < 2.2 Å) to obtain reliable protein:ligand bond distances, to unambiguously assign rotameric conformations of side chains and ligand functions, and to visualize water molecules? Is the active site unoccupied, or are there naturally occurring ligands or other nonwater molecules residing within it that may hamper ligand-soaking experiments? Is the active site open to solvent channels in the crystal or is it obstructed by crystal packing? Is the active site conformationally restricted or otherwise rendered incapable of undergoing structural rearrangements necessary for ligand binding? In situations where such pathologies do exist, the only solution may be to engineer a new variant and screen for a new crystal form, as was demonstrated for urokinase inhibitor design discussed earlier in this chapter (Section 3.38.3.2.3.3). Barring any of the aforementioned problems, the crystallographer can pursue a number of integrated strategies to generate and harvest the highly valued structural data. At this stage, an integrated effort between the crystallographers, the modelers, and the medicinal chemists cannot be overemphasized, as an early collaboration will undoubtedly speed the progression from hits to leads, and from leads to compounds with therapeutic potential (Figure 7).

3.38.3.4.1 Crystallography-driven lead characterization: the traditional view

The conventional approach to lead discovery involves screening the target against a high-throughput screening (HTS) library (see 3.32 High-Throughput and High-Content Screening). The HTS library can contain upwards of 2 million compounds and therefore implicitly requires high-throughput methods to make the time scale of the process manageable. A biochemical assay amenable to high throughput is normally devised for the screening, and compounds with an IC_{50} of ≤ 20 μ M are typically selected for further testing (see 3.27 Optical Assays in Drug Discovery; 3.28 Fluorescence Screening Assays; 3.29 Cell-Based Screening Assays). These ‘hits’ may be handed over to medicinal chemists for optimization of potency or other druglike properties, or delivered to the crystallographers for structural characterization of the target:ligand complex. Given the resources, both investigatory avenues might be pursued concurrently, and it is therefore vital for the crystallographers to keep pace with the medicinal chemists. The iterative nature of SBDD, in that each cycle of design, synthesis, and bioassay is followed by the crystal structure determination of the target:ligand complex, requires that the individual disciplines work effectively with one another (see 4.24 Structure-Based Drug Design – The Use of Protein Structure in Drug Discovery). This iterative approach can be applied to SBDD in several ways. A powerful technique is de novo drug design – the design of novel chemical structures from scratch – as guided by the crystal structure of an empty active site (see 4.13 De Novo Design). Another strategy is lead optimization, which begins with the crystal structure of a target:lead complex. The lead may have been discovered through the chemical or patent literature, through HTS of compound libraries, or sourced from de novo design. Once the complex of the lead and target protein has been elucidated crystallographically, the lead can be modified in order to optimize its chemical or biological properties (see 5.01 The Why and How of Absorption, Distribution, Metabolism, Excretion, and Toxicity Research).

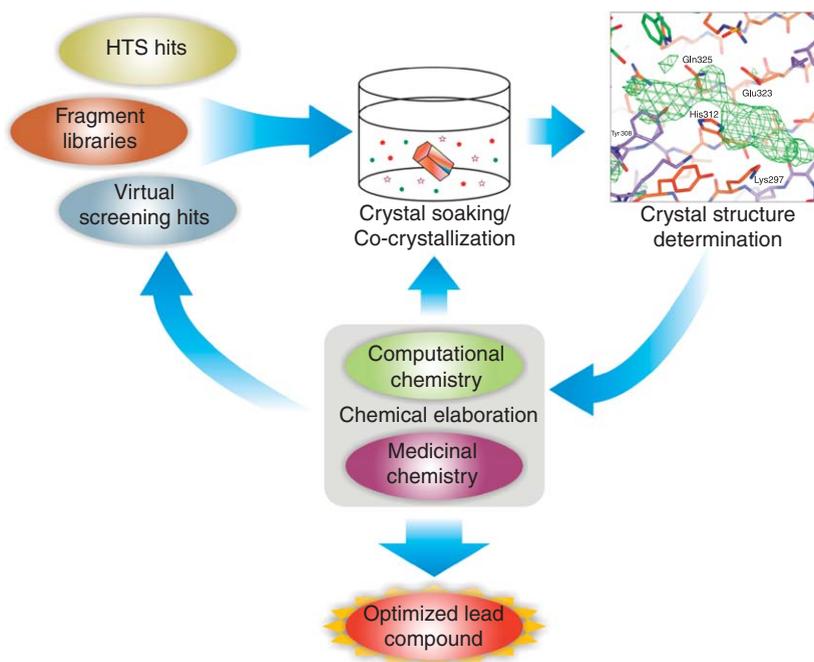


Figure 7 Flowpath illustrating the integrated approach to crystallography-driven SBDD. Small molecule (“fragment”) libraries or promising compounds identified from conventional HTS or virtual screening are subjected to co-crystallization or crystal-soaking trials with the target macromolecule (see 3.32 High-Throughput and High-Content Screening; 3.41 Fragment-Based Approaches; 4.19 Virtual Screening). Crystal structure determination of the target:ligand complex provides critical data for computational exploration and medicinal chemistry decisions. Characterization of the target:ligand binding interactions may also guide selective screening of focused libraries, which can produce new hits for the SBDD cycle. Several iterative cycles of this process may lead to an optimized lead compound.

For the crystallographer, a fundamental component of the design cycle is successful structure determination of the complex to high resolution. Complexes are obtained by target:ligand co-crystallization or by diffusing the ligand into preformed crystals (soaking). Both methods have their particular strengths and weaknesses. Soaking takes advantage of the fact that macromolecular crystals are solvent rich¹⁹⁵ and contain integral networks of solvent channels; however, the size and configuration of the channels within the crystal lattice will place an upper limit on the size of ligands that may be diffused in (this is usually not an issue for small druglike molecules). As mentioned, crystal lattice restraints can sometimes hinder conformational rearrangements of the target, which might necessarily coincide with ligand binding or, on the other hand, any rearrangements that do take place may disrupt the crystal lattice, causing irreversible damage. With the co-crystallization approach the target:ligand complex is formed before crystallization is carried out, but any alterations in the physicochemical properties of the complex relative to the native protein may preclude crystallization of the complex under native conditions. Herein lie some of the perceived advantages of the soaking method: native crystals can be grown reproducibly, and the conservation of isomorphism between native and soaked crystals allows for rapid detection of the bound ligand by difference Fourier methods. Does this mean that co-crystallization methods are less amenable to high-throughput SBDD? On the contrary, the Plexxikon group has recently demonstrated the effectiveness of a co-crystallization approach in their SBDD program targeting human phosphodiesterases.¹⁹⁶

3.38.3.4.1.1 Ligand soaking and co-crystallization: practical considerations

In order for a compound to be of value to SBDD, it must be present at a high enough concentration in the crystallization drop to promote a high occupancy of binding to the target and therefore produce interpretable electron density. A source of difficulty can be limited ligand solubility, which can impose a constraint on binding occupancy and limit the usefulness of the crystallographic data. If something is known about the target:ligand dissociation constant, K_D , the problem can be approached using the formula $[\text{ligand}] = [\text{protein}] + nK_D$, where $[\text{ligand}]$ and $[\text{protein}]$ are the concentrations in the crystallization drop, and n is an ‘excess factor.’ When $K_D \ll [\text{protein}]$, for example, when a ligand with a K_D in the low micromolar range is added to a drop containing a millimolar concentration of protein, then

[ligand] \cong [protein]. In this case, maximum binding occupancy should be achieved by adding only a slight molar excess of ligand. On the other hand, when the protein is crystallized at low concentration and/or a low affinity ligand is used (i.e., $K_D \cong$ [protein]), then the nK_D term becomes appreciable. A general guideline is to use an n value of at least 5–10, although trial-and-error comes into play here. An added complication is that the above calculation relies on the assumption that the measured K_D under assay conditions is similar to the K_D in the crystallization drop, which is very likely untrue. For a polar ligand, high salt concentrations will normally raise the K_D in the crystallization drop, while the opposite will be true for nonpolar ligands. Potentially adding to the puzzle is the effect of cryoprotectants and cosolvents, such as polyethylene glycol, which have been known to sometimes interfere with ligand binding by competing for the ligand-binding site. All considered, a ‘more is better’ philosophy for ligand soaking is often adhered to, however poorly soluble ligands may require the use of a suitable organic solvent such as dimethyl sulfoxide (DMSO), ethanol, methanol, or hexafluoropropanol.¹⁸⁰

3.38.3.4.2 Crystallography and ‘fragonomics’: the emerging area of fragment-based drug design

Novel approaches to drug discovery have been developed recently that utilize biophysical methods to screen collections of basic chemical building blocks, termed ‘fragments’ (see 3.41 Fragment-Based Approaches).^{197,198} Fragment libraries tend to contain ~100–1000 compounds and are therefore much smaller than conventional HTS libraries. The compounds themselves are also smaller and simpler (average molecular weight 100–250 Da; 8–18 nonhydrogen atoms) than typical HTS compounds and, consequently, fragments bind to target proteins weakly with an affinity that is below the detection limit of conventional HTS methods. Biophysical techniques such as NMR¹⁹⁹ (see 3.39 Nuclear Magnetic Resonance in Drug Discovery) or x-ray crystallography,^{200–203} however, are powerful screening methods for fragment-based searches. It is now generally accepted that target:fragment interactions as low as millimolar K_D can be reliably detected using NMR or x-ray crystallography. X-ray-based screening has the added advantage of inherently producing high-resolution structural data that can provide a basis for computational design and medicinal chemistry decisions. But what constitutes a good fragment library? An analysis of fragment hits against a variety of targets has led the Astex group to formulate a ‘rule of three’²⁰⁴ for describing a typical fragment ‘hit’: molecular weight ≤ 300 Da; hydrogen bond donors ≤ 3 ; hydrogen bond acceptors ≤ 3 ; octanol/water partition coefficient ($\log P$) ≤ 3 ; and number of rotatable bonds ≤ 3 .

The first well-publicized crystal screening method, termed CrystalLEAD, was developed in the Abbott labs, and the practical application of the method was initially demonstrated on the anticancer target urokinase.^{200,205} In this study, shape-diverse compound mixtures were soaked into crystals of an engineered form of the protein, termed micro-urokinase (Section 3.38.3.2.3.3). Four mixtures yielded evidence of binding by five novel ligands, with one mixture containing two binders where the less potent ligand was only discovered after rescreening with the most potent inhibitor removed. First-round optimization of one of the leads, 8-hydroxy-2-aminoquinoline, to 8-aminopyrimidyl-2-aminoquinoline, resulted in a 100-fold increase in inhibitor potency (K_i from 56 μ M to 370 nM), and provided a favorable starting point for further lead development.

3.38.3.4.2.1 ‘Scaffold-based’ discovery and tethering: variations on a theme

The demonstrated success of fragment-based methods has bolstered further exploration and development of the approach. The Plexxikon group has recently introduced a crystallography-driven screening technique called ‘scaffold-based’ drug design, which they implemented in their program to develop novel inhibitors of human phosphodiesterases (PDEs).¹⁹⁶ According to the Plexxikon definition, the main difference between libraries containing molecular scaffolds vs. molecular fragments is that the former comprises slightly larger molecules (average molecular weight ~250 Da) containing additional functional groups that enable them to bind to the target with an affinity detectable by conventional HTS. With this approach, a preliminary HTS assay is used at a compound concentration of ~200 μ M for the initial selection of candidates. The target protein is then crystallized in the presence of 1 mM of the scaffold molecule, with ranges of pH, precipitants, and additives screened to help increase the chances of co-crystallization. The authors reported an overall co-crystallization success rate of 85% for the 316 chemical scaffolds selected for structural studies with different PDEs.¹⁹⁶

Scientists at Sunesis have developed a proprietary fragment-based method, called Tethering, which offers a site-directed means to probe ligand binding to a target. Tethering takes advantage of native or engineered surface-exposed cysteines on the target to capture thiol-containing ligands by formation of a mixed disulfide.^{206–208} Crystallographic analysis of the covalently linked complex provides a powerful means for rapid identification of fragments that can be converted into reversible inhibitors, as exemplified by the group’s development of novel allosteric and active-site caspase inhibitors.^{209,210}

3.38.3.4.2.2 Lead identification

Once the interactions between a fragment and its target have been characterized by crystallographic methods, iterative structure-driven chemistry can be applied to generate a drug-size lead compound (*see* 3.41 Fragment-Based Approaches). These subsequent steps can be subdivided into four distinct categories: fragment optimization; fragment linking; fragment evolution; and fragment self-assembly¹⁹⁸ (Figure 8). In fragment evolution, a lead molecule is iteratively generated, or ‘evolved’, by building away from the starting fragment into other regions of the binding pocket. Fragment linking defines the process of chemically joining two fragments that bind adjacent sites to generate a larger compound with greater potency. Fragment optimization involves altering properties of a lead molecule including selectivity or efficacy, such as a ‘SAR by x-ray’ approach.²¹¹ The method of fragment self-assembly utilizes mixtures of fragments adorned with reactive groups (so-called ‘dynamic combinatorial libraries’²¹²), which link together when situated in close proximity (e.g., within a target’s active site) (*see* 3.42 Dynamic Ligand Assembly). In this way, the target catalyzes the synthesis of its own – and presumably highest affinity – lead from a library of fragments. An x-ray-guided extension to fragment self-assembly, termed ‘dynamic combinatorial x-ray crystallography’ (DCX), allows self-assembled leads to be generated from dynamic combinatorial libraries in the presence of protein crystals, permitting direct observation of the binding interaction with the target.²¹³

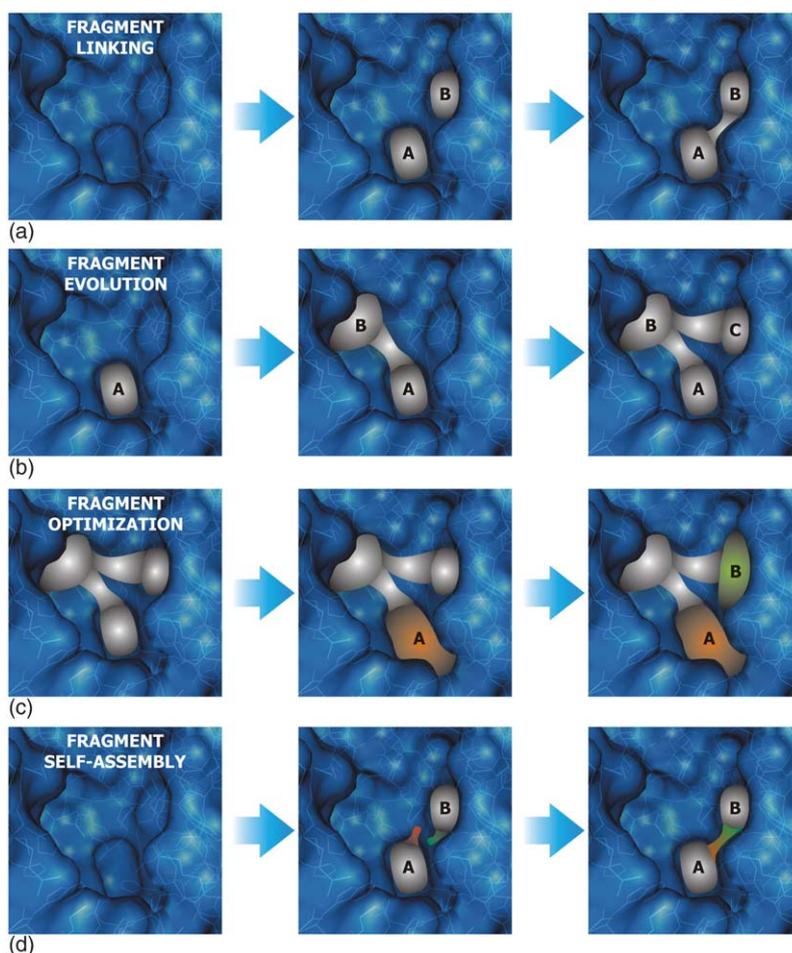


Figure 8 Fragment-based lead generation and optimization approaches. (a) Fragment linking involves the joining of fragments identified in vicinal binding sites with a chemical linker, leading to a larger molecule with higher binding affinity. (b) Fragment evolution proceeds by iteratively building out from a single starting fragment to create a larger, more complex molecule that interacts with neighboring pockets. (c) Fragment optimization entails the rational re-engineering of a lead molecule in order to improve or alter specific properties of the molecule (e.g., affinity, selectivity, oral bioavailability). (d) Fragment self-assembly exploits the ability of reactive fragments with complementary functional groups to assemble into a larger, more potent molecule when situated in proximal binding pockets on a target template.

3.38.4 The Era of High-Throughput Crystallography in Lead Discovery

In the past, the determination of a protein x-ray crystal structure typically required several years and, in difficult cases, even decades. The main bottleneck of obtaining suitable amounts of pure protein samples was drastically reduced with the advent of molecular biology tools developed in the 1980s and 1990s (see 3.19 Protein Production for Three-Dimensional Structural Analysis). The structure determination process, which has traditionally been a lengthy and laborious undertaking, has greatly benefited from sweeping technological and methodological breakthroughs. Advances in task automation, miniaturization, and parallelization have ushered in a new era of 'high-throughput crystallography.' The original notion of high-throughput crystallography was not conceived by the pharmaceutical sector (which had diminishing interests in x-ray crystallography and was instead focusing on other HT technologies), but by the structural genomics initiatives that were established with the goal of solving crystal structures of representatives from all known protein families (see 3.25 Structural Genomics).^{214,215} Such ambitious goals bore the immediate need to establish new methodologies for producing hundreds of proteins in parallel, for creating HT systems to dispense and monitor thousands of crystallization experiments per day, and for determining the resultant x-ray structures in a fully automated fashion. The technological advances that have emerged from the various structural genomics initiatives have not only helped to improve our understanding of protein fold space, but has also substantially benefited the pharmaceutical industry by adding x-ray crystallography to a preexisting armamentarium of HT tools for drug discovery.

For the design of DNA constructs, a genomics approach is usually taken at the outset: bioinformatics tools are used to mine available databases for homologous sequences from various organisms, and multiple-sequence alignments are examined to predict a possible domain structure for the protein (see 3.15 Bioinformatics). Several orthologs are generated with different start/stop points delineating the domain(s) of interest, and appropriate oligonucleotide primers are either synthesized or ordered en masse. Microtiter plates (typically 96-well or higher) are used for automated HT-PCR, cloning, and testing of expression.^{216–220} The attainability of large-scale automated production and purification of proteins has been demonstrated in the construction of an elaborate robotics system capable of handling 96 parallel 65–70 ml bacterial cultures with average yields of 10 mg of purified protein per sample.¹⁴ Bacterial cell-free systems, such as those based on the *E. coli* S30 extract, offer opportunities for automation of the production of milligram quantities of protein in a test tube.^{221,222}

Robotic systems for automated crystallization with smaller volumes and higher density plating configurations (96-, 384-, and 1536-well formats), along with imaging systems for automated crystallization drop monitoring and scoring, have enabled HT crystallization on a miniaturized scale and resulted in drastic reductions in space, material, and cost.^{223–225} Hardware and software solutions for automated crystal mounting, alignment, and data collection can allow rapid, unmanned data collection for synchrotron^{226–229} and laboratory^{230,231} sources. Software developments for the automation of macromolecular structure determination are enabling a rapid and seamless transfer of diffraction data through the processing, phasing, and model-building pipeline.^{232–237} Adding to this are new programs that provide solutions for automatic generation of ligand coordinates and topologies²³⁸ and automated ligand building into electron-density maps.^{239–243} Continued development of HT crystallography hardware and software applications and a global integration of the individual processes will further hasten the lead identification and optimization cycle.

References

1. Fleming, A. B. *J. Exp. Pathol.* **1929**, *10*, 226–236.
2. Hillisch, A.; Hilgenfeld, R. *Modern Methods of Drug Discovery*; Birkhäuser Verlag: Basel, Switzerland, 2003.
3. Hillisch, A.; Pineda, L. F.; Hilgenfeld, R. *Drug Disc. Today* **2004**, *9*, 659–669.
4. Stevens, R. C. *Drug Disc. World* **2003**, *4*, 35–48.
5. Steitz, T. A.; Ludwig, M. L.; Quijcho, F. A.; Lipscomb, W. N. *J. Biol. Chem.* **1967**, *242*, 4662–4668.
6. Lipscomb, W. N.; Hartsuck, J. A.; Quijcho, F. A.; Reeke, G. N., Jr. *Proc. Natl. Acad. Sci. USA* **1969**, *64*, 28–35.
7. Quijcho, F. A.; Lipscomb, W. N. *Adv. Prot. Chem.* **1971**, *25*, 1–78.
8. Byers, L. D.; Wolfenden, R. *Biochemistry* **1973**, *12*, 2070–2078.
9. Ondetti, M. A.; Rubin, B.; Cushman, D. W. *Science* **1977**, *196*, 441–444.
10. Natesh, R.; Schwager, S. L. U.; Sturrock, E. D.; Acharya, K. R. *Nature* **2003**, *421*, 551–554.
11. Natesh, R.; Schwager, S. L. U.; Evans, H. R.; Sturrock, E. D.; Acharya, K. R. *Biochemistry* **2004**, *43*, 8718–8724.
12. Stevens, R. C. *Structure* **2000**, *8*, R177–R185.
13. Gilbert, M.; Albala, J. S. *Curr. Opin. Chem. Biol.* **2001**, *6*, 102–105.
14. Lesley, S. A. *Protein Exp. Purif.* **2001**, *22*, 159–164.
15. Nagar, B.; Hantschel, O.; Young, M. A.; Scheffzek, K.; Veach, D.; Bornmann, W.; Clarkson, B.; Superti-Furga, G.; Kuriyan, J. *Cell* **2003**, *112*, 859–871.
16. Hogg, T.; Mechold, M.; Malke, H.; Cashel, M.; Hilgenfeld, R. *Cell* **2004**, *117*, 57–68.
17. Gao, X.; Bain, K.; Bonanno, J. B.; Buchanan, M.; Henderson, D.; Lorimer, D.; Marsh, C.; Reynes, J. A.; Sauder, J. M.; Schwinn, K.; Thai, C.; Burley, S. K. *J. Struct. Funct. Genomics* **2005**, *6*, 129–134.

18. Prilusky, J.; Felder, C. E.; Zeev-Ben-Mordechai, T.; Rydberg, E. H.; Man, O.; Beckmann, J. S.; Silman, I.; Sussman, J. L. *Bioinformatics* **2005**, *21*, 3435–3438.
19. Linding, R.; Jensen, L. J.; Diella, F.; Bork, P.; Gibson, T. J.; Russell, R. B. *Structure* **2003a**, *11*, 1453–1459.
20. Ward, J. J.; Sodhi, J. S.; McGuffin, L. J.; Buxton, B. F.; Jones, D. T. *J. Mol. Biol.* **2004**, *337*, 635–645.
21. Linding, R.; Russell, R. B.; Neduva, V.; Gibson, T. J. *Nucleic Acids Res.* **2003b**, *31*, 3701–3708.
22. Dosztanyi, Z.; Csizsmok, V.; Tompa, P.; Simon, I. *J. Mol. Biol.* **2005**, *347*, 827–839.
23. Romero, P.; Obradovic, Z.; Li, X.; Garner, E. C.; Brown, C. J.; Dunker, A. K. *Proteins* **2001**, *42*, 38–48.
24. Coezytaux, K.; Poupon, A. *Bioinformatics* **2005**, *21*, 1891–1900.
25. Yang, Z. R.; Thomson, R.; McNeil, P.; Esnouf, R. M. *Bioinformatics* **2005**, *21*, 3369–3376.
26. Vucetic, S.; Brown, C. J.; Dunker, A. K.; Obradovic, Z. *Proteins* **2003**, *52*, 573–584.
27. Peng, K.; Vucetic, S.; Radivojac, P.; Brown, C. J.; Dunker, A. K.; Obradovic, Z. *J. Bioinform. Comput. Biol.* **2005**, *3*, 35–60.
28. Moulton, J. *Curr. Opin. Struct. Biol.* **2005**, *15*, 285–289.
29. Vucetic, S.; Obradovic, Z.; Vacic, V.; Radivojac, P.; Peng, K.; Iakoucheva, L. M.; Cortese, M. S.; Lawson, J. D.; Brown, C. J.; Sikes, J. G.; Newton, C. D.; Dunker, A. K. *Bioinformatics* **2005**, *21*, 137–140.
30. Englander, J. J.; Del Mar, C.; Li, W.; Englander, S. W.; Kim, J. S.; Stranz, D. D.; Hamuro, Y.; Woods, V. L., Jr. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 7057–7062.
31. Pantazatos, D.; Kim, J. S.; Klock, H. E.; Stevens, R. C.; Wilson, I. A.; Lesley, S. A.; Woods, V. L., Jr. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 751–756.
32. Bornhorst, J. A.; Falke, J. J. *Methods Enzymol.* **2000**, *326*, 245–254.
33. Lichty, J. J.; Malecki, J. L.; Agnew, H. D.; Michelson-Horowitz, D. J.; Tan, S. *Protein Expr. Purif.* **2005**, *41*, 98–105.
34. Sachdev, D.; Chirgwin, J. M. *Methods Enzymol.* **2000**, *326*, 312–321.
35. LaVallie, E. R.; Lu, Z.; Diblasio-Smith, E. A.; Collins-Racie, L. A.; McCoy, J. M. *Methods Enzymol.* **2000**, *326*, 322–340.
36. Smith, D. B. *Methods Enzymol.* **2000**, *326*, 254–270.
37. Smyth, D. R.; Mrozkiewicz, M. K.; McGrath, W. J.; Listwan, P.; Kobe, B. *Protein Sci.* **2003**, *12*, 1313–1322.
38. Yee, A.; Chang, X.; Pineda-Lucena, A.; Wu, B.; Semesi, A.; Le, B.; Ramelot, T.; Lee, G. M.; Bhattacharyya, S.; Gutierrez, P. et al. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 1825–1830.
39. Morton, C. L.; Potter, P. M. *Mol. Biotechnol.* **2000**, *16*, 193–202.
40. Li, P.; Gao, X. G.; Arellano, R. O.; Renugopalakrishnan, V. *Protein Expr. Purif.* **2001**, *22*, 369–380.
41. Hamilton, S. R.; Bobrowicz, P.; Bobrowicz, B.; Davidson, R. C.; Li, H.; Mitchell, T.; Nett, J. H.; Rausch, S.; Stadheim, T. A.; Wischniewski, H. et al. *Science* **2003**, *301*, 1244–1246.
42. Sulzenbacher, G.; Gruez, A.; Roig-Zamboni, V.; Spinelli, S.; Valencia, C.; Pagot, F.; Vincentelli, R.; Bignon, C.; Salomoni, A.; Grisel, S. et al. *Acta Crystallogr.* **2002**, *D58*, 2109–2115.
43. Clark, E. D. B. *Curr. Opin. Biotechnol.* **1998**, *9*, 157–163.
44. Armstrong, N.; de Lencastre, A.; Gouaux, E. *Protein Sci.* **1999**, *8*, 1475–1483.
45. Chen, X. S.; Stehle, T.; Harrison, S. C. *EMBO J.* **1998**, *17*, 3233–3240.
46. Hofmeister, F. *Arch. Exp. Pathol. Pharmacol. (Leipzig)* **1888**, *24*, 247–260.
47. Kunz, W.; Henle, J.; Ninham, B. W. *Curr. Opin. Coll. Interface Sci.* **2004**, *9*, 19–37.
48. Arakawa, T.; Timasheff, S. N. *Methods Enzymol.* **1985**, *114*, 49–77.
49. Arakawa, T.; Bhat, R.; Timasheff, S. N. *Biochemistry* **1990**, *29*, 1924–1931.
50. Gekko, K.; Timasheff, S. N. *Biochemistry* **1981**, *20*, 4677–4686.
51. Womack, M. D.; Kendall, D. A.; MacDonald, R. C. *Biochim. Biophys. Acta* **1983**, *733*, 210–215.
52. Asakura, T.; Adachi, K.; Schwartz, E. J. *Biol. Chem.* **1978**, *253*, 6423–6425.
53. Hedman, P. O.; Gustafsson, J. G. *Anal. Biochem.* **1984**, *138*, 411–415.
54. Gerlsma, S. Y. J. *Biol. Chem.* **1968**, *243*, 957–961.
55. Lee, J. C.; Timasheff, S. N. *J. Biol. Chem.* **1981**, *256*, 7193–7201.
56. Arakawa, T.; Timasheff, S. N. *Biochemistry* **1982**, *21*, 6536–6544.
57. Lindwall, G.; Chau, M.-F.; Gardner, S. R.; Kohlstaedt, L. A. *Protein Eng.* **2000**, *13*, 67–71.
58. Jancarik, J.; Pufan, R.; Hong, C.; Kim, S. H.; Kim, R. *Acta Crystallogr.* **2004**, *D60*, 1670–1673.
59. Collins, B. K.; Tomanicek, S. J.; Lyamicheva, N.; Kaiser, M. W.; Mueser, T. C. *Acta Crystallogr.* **2004**, *D60*, 1674–1678.
60. Hünefeld, F. L.; *Der Chemismus in der tierischen Organisation*; FA. Brockhaus: Leipzig, Germany, 1840, p 160.
61. McPherson, A. *Crystallization of Biological Macromolecules*; CSHL Press: New York, 1999, pp 1–10.
62. Hogg, T.; Smatanova, I. K.; Bezouska, K.; Ulbrich, N.; Hilgenfeld, R. *Acta Crystallogr.* **2002**, *D58*, 1734–1739.
63. Gevaert, K.; Vandekerckhove, J. *Electrophoresis* **2000**, *21*, 1145–1154.
64. Yates, J. R., III. *Trends Genet.* **2000**, *16*, 5–8.
65. Yates, J. R., III. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33*, 297–316.
66. Ferre-D'Amare, A. R.; Burley, S. K. *Structure* **1994**, *2*, 357–359.
67. D'Arcy, A. *Acta Crystallogr.* **1994**, *D50*, 469–471.
68. Ferre-D'Amare, A. R.; Burley, S. K. *Methods Enzymol.* **1997**, *276*, 157–166.
69. Zulauf, M.; D'Arcy, A. *J. Cryst. Growth* **1992**, *122*, 102–106.
70. Han, K. K.; Martinage, A. *Int. J. Biochem.* **1992**, *24*, 19–28.
71. Baker, H. M.; Day, C. L.; Norris, G. E.; Baker, E. N. *Acta Crystallogr.* **1994**, *D50*, 380–384.
72. Grueninger-Leitch, F.; D'Arcy, A.; D'Arcy, B.; Chène, C. *Protein Sci.* **1996**, *5*, 2617–2622.
73. Hoover, D. M.; Schalk-Hihi, C.; Chou, C.; Menon, S.; Wlodawer, A.; Zdanov, A. *Eur. J. Biochem.* **1999**, *262*, 134–141.
74. Dale, G. E.; D'Arcy, B.; Yuvaniyama, C.; Wipf, B.; Oefner, C.; D'Arcy, A. *Acta Crystallogr.* **2000**, *D56*, 894–897.
75. Davis, S. J.; Ikemizu, S.; Collins, A. V.; Fennelly, J. A.; Harlos, K.; Jones, E. Y.; Stuart, D. I. *Acta Crystallogr.* **2001**, *D57*, 605–608.
76. Tarentino, A. L.; Plummer, T. H., Jr. *Methods Enzymol.* **1994**, *230*, 44–57.
77. Iwase, H.; Hotta, K. *Methods Mol. Biol.* **1993**, *14*, 151–159.
78. Uchida, Y.; Tsukada, Y.; Sugimori, T. *J. Biochem. (Tokyo)* **1979**, *86*, 1573–1585.
79. Binns, K. L.; Taylor, P. P.; Sicheri, F.; Pawson, T.; Holland, S. J. *Mol. Cell. Biol.* **2000**, *20*, 4791–4805.
80. Yang, J.; Cron, P.; Good, V. M.; Thompson, V.; Hemmings, B. A.; Barford, D. *Nat. Struct. Biol.* **2002**, *9*, 940–944.

81. Yang, J.; Cron, P.; Thompson, V.; Good, V. M.; Hess, D.; Hemmings, B. A.; Barford, D. *Mol. Cell.* **2002**, *9*, 1227–1240.
82. Prongay, A. J.; Smith, T. J.; Rossmann, M. G.; Ehrlich, L. S.; Carter, C. A.; McClure, J. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 9980–9984.
83. Yu, X. C.; Sturrock, E. D.; Wu, Z.; Biemann, K.; Ehlers, M. R.; Riordan, J. F. *J. Biol. Chem.* **1997**, *272*, 3511–3519.
84. Gordon, K.; Redelinghuys, P.; Schwager, S. L.; Ehlers, M. R.; Papageorgiou, A. C.; Natesh, R.; Acharya, K. R.; Sturrock, E. D. *Biochem. J.* **2003**, *371*, 437–442.
85. Nienaber, V.; Wang, J.; Davidson, D.; Henkin, J. *J. Biol. Chem.* **2000**, *275*, 7239–7248.
86. Kimber, M. S.; Vallee, F.; Houston, S.; Nečakov, A.; Skarina, T.; Evdokimova, E.; Beasley, S.; Christendat, D.; Savchenko, A.; Arrowsmith, C. H. et al. *Proteins* **2003**, *51*, 562–568.
87. Page, R.; Grzechnik, S. K.; Canaves, J. M.; Spraggon, G.; Kreusch, A.; Kuhn, P.; Stevens, R. C.; Lesley, S. A. *Acta Crystallogr.* **2003**, *D59*, 1028–1037.
88. Rupp, B. *J. Struct. Biol.* **2003**, *142*, 162–169.
89. Kwong, P. D.; Wyatt, R.; Desjardins, E.; Robinson, J.; Culp, J. S.; Hellmig, B. D.; Sweet, R. W.; Sodroski, J.; Hendrickson, W. A. *J. Biol. Chem.* **1999**, *274*, 4115–4123.
90. Kwong, P. D.; Wyatt, R.; Robinson, J.; Sweet, R. W.; Sodroski, J.; Hendrickson, W. A. *Nature* **1998**, *393*, 648–659.
91. Kendrew, J. C.; Parrish, R. G.; Marrack, J. R.; Orlans, E. S. *Nature* **1954**, *174*, 946–949.
92. Kendrew, J. C.; Parrish, R. G. *Proc. R. Soc. Lond. A* **1956**, *238*, 305–324.
93. Campbell, J. W.; Duee, E.; Hodgson, G.; Mercer, W. D.; Stammers, D. K.; Wendell, P. L.; Muirhead, H.; Watson, H. C. *Cold Spring Harbor Symp. Quant. Biol.* **1972**, *36*, 165–170.
94. Nikolov, D. B.; Hu, S. H.; Lin, J.; Gasch, A.; Hoffmann, A.; Horikoshi, M.; Chua, N. H.; Roeder, R. G.; Burley, S. K. *Nature* **1992**, *360*, 40–46.
95. Savchenko, A.; Yee, A.; Khachatryan, A.; Skarina, T.; Evdokimova, E.; Pavlova, M.; Semesi, A.; Northey, J.; Beasley, S.; Lan, N. et al. *Proteins* **2003**, *50*, 392–399.
96. Appelt, K.; Bacquet, R. J.; Bartlett, C. A.; Boothe, C. L. J.; Freer, S. T.; Fuhry, M. M.; Gehring, M. R.; Hermann, S. M.; Howland, E. F.; Janson, C. A. et al. *J. Med. Chem.* **1991**, *34*, 1925–1934.
97. Varney, M. D.; Marzoni, G. P.; Palmer, C. L.; Deal, J. G.; Webber, S.; Welsh, K. M.; Bacquet, R. J.; Bartlett, C. A.; Morse, C. A.; Booth, C. L. J. et al. *J. Med. Chem.* **1992**, *35*, 663–676.
98. Webber, S. E.; Bleckman, T. M.; Attard, J.; Deal, J. G.; Kathardekar, V.; Welsh, K. M.; Webber, S.; Janson, C. A.; Matthews, D. A.; Smith, W. W. et al. *J. Med. Chem.* **1993**, *36*, 733–746.
99. Ikuta, M.; Kamata, K.; Fukasawa, K.; Honma, T.; Machida, T.; Hirai, H.; Suzuki-Takahashi, I.; Hayama, T.; Nishimura, S. *J. Biol. Chem.* **2001**, *276*, 27548–27554.
100. Anand, K.; Ziebuhr, J.; Wadhvani, P.; Mesters, J. R.; Hilgenfeld, R. *Science* **2003**, *300*, 1763–1767.
101. Keenan, R. J.; Siehl, D. L.; Gorton, R.; Castle, L. A. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 8887–8892.
102. Lawson, D. M.; Artymiuk, P. J.; Yewdall, S. J.; Smith, J. M. A.; Livingstone, J. C.; Treffry, A.; Luzzago, A.; Levi, S.; Arosio, P.; Cesareni, G. et al. *Nature* **1991**, *349*, 541–544.
103. Laufferberger, V. *Bull. Soc. Chim. Biol.* **1937**, *19*, 1575–1582.
104. Banyard, S. H.; Stammers, D. K.; Harrison, P. M. *Nature* **1978**, *271*, 282–284.
105. Ford, G. C.; Harrison, P. M.; Rice, D. W.; Smith, J. M.; Treffry, A.; White, J. L.; Yariv, J. *Philos. Trans. R. Soc. Lond. Ser. B* **1984**, *304*, 551–556.
106. Thomas, C. D.; Shaw, W. V.; Lawson, D. M.; Treffry, A.; Artymiuk, P. J.; Harrison, P. M. *Biochem. Soc. Trans.* **1988**, *16*, 838–839.
107. Arosio, P.; Gatti, G.; Bolognesi, M. *Biochim. Biophys. Acta* **1983**, *744*, 230–232.
108. Abel, J. J. *Proc. Natl. Acad. Sci. USA* **1926**, *12*, 132–136.
109. Riboldi-Tunnicliffe, A.; Konig, B.; Jessen, S.; Weiss, M. S.; Rahfeld, J.; Hacker, J.; Fischer, G.; Hilgenfeld, R. *Nat. Struct. Biol.* **2001**, *8*, 779–783.
110. Trakhanov, S.; Quioco, F. A. *Protein Sci.* **1995**, *4*, 1914–1919.
111. Trakhanov, S.; Kreimer, D. I.; Parkin, S.; Ames, G. F.; Rupp, B. *Protein Sci.* **1998**, *7*, 600–604.
112. Harding, M. M. *Acta Crystallogr.* **2000**, *D56*, 857–867.
113. Harding, M. M. *Acta Crystallogr.* **2001**, *D57*, 401–411.
114. Hendrickson, W. A. *Science* **1991**, *254*, 51–58.
115. Qiu, X.; Janson, C. A. *Acta Crystallogr.* **2004**, *D60*, 1545–1554.
116. Dale, G. E.; Oefner, C.; D'Arcy, A. *J. Struct. Biol.* **2003**, *142*, 88–97.
117. Derewenda, Z. S. *Structure* **2004**, *12*, 529–535.
118. Braig, K.; Otwinowski, Z.; Hegde, R.; Boisvert, D. C.; Joachimiak, A.; Horwich, A. L.; Sigler, P. B. *Nature* **1994**, *371*, 578–586.
119. Horwich, A. *Nat. Struct. Biol.* **2000**, *7*, 269–270.
120. McElroy, H. E.; Sisson, G. W.; Schoettlin, W. E.; Aust, R. M.; Villafranca, J. E. *J. Cryst. Growth* **1992**, *122*, 265–272.
121. D'Arcy, A.; Stihle, M.; Kostrewa, D. A.; Dale, G. *Acta Crystallogr.* **1999**, *D55*, 1623–1625.
122. Dyda, F.; Hickman, A. B.; Jenkins, T. M.; Engelman, A.; Craigie, R.; Davies, D. R. *Science* **1994**, *266*, 1981–1986.
123. Fallner, M.; Niederweis, M.; Schulz, G. E. *Science* **2004**, *303*, 1189–1192.
124. Zhang, F.; Basinski, M. B.; Beals, J. M.; Briggs, S. L.; Churgay, L. M.; Clawson, D. K.; DiMarchi, R. D.; Furman, T. C.; Hale, J. E.; Hsiung, H. M. et al. *Nature* **1997**, *387*, 206–209.
125. Schlatter, D.; Thoma, R.; Küng, E.; Stihle, M.; Müller, F.; Borroni, E.; Cesura, A.; Hennig, M. *Acta Crystallogr.* **2005**, *D61*, 513–519.
126. Derewenda, Z. S. *Methods* **2004**, *34*, 354–363.
127. Baud, F.; Karlin, S. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 12494–12499.
128. Avbelj, F.; Fele, L. *J. Mol. Biol.* **1998**, *279*, 665–684.
129. Lo Conte, L.; Chothia, C.; Janin, J. *J. Mol. Biol.* **1999**, *285*, 2177–2198.
130. Longenecker, K. L.; Garrard, S. M.; Sheffield, P. J.; Derewenda, Z. S. *Acta Crystallogr.* **2001**, *D57*, 679–688.
131. Mateja, A.; Devedjiev, Y.; Krowarsch, D.; Longenecker, K.; Dauter, Z.; Otlewski, J.; Derewenda, Z. S. *Acta Crystallogr.* **2002**, *D58*, 1983–1991.
132. Munshi, S.; Hall, D. L.; Kornienko, M.; Darke, P. L.; Kuo, L. C. *Acta Crystallogr.* **2003**, *D59*, 1725–1730.
133. Dasgupta, S.; Iyer, G. H.; Bryant, S. H.; Lawrence, C. E.; Bell, J. A. *Proteins* **1997**, *28*, 494–514.
134. Czepas, J.; Devedjiev, Y.; Krowarsch, D.; Derewenda, U.; Otlewski, J.; Derewenda, Z. S. *Acta Crystallogr.* **2004**, *D60*, 275–280.
135. Wingren, C.; Edmundson, A. B.; Borrebaeck, C. A. K. *Protein Eng.* **2003**, *16*, 255–264.
136. Waldo, G. S.; Standish, B. M.; Berendzen, J.; Terwilliger, T. C. *Nat. Biotechnol.* **1999**, *17*, 691–695.
137. Yang, J. K.; Park, M. S.; Waldo, G. S.; Suh, S. W. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 455–460.

138. Rayment, I.; Rypniewski, W. R.; Schmidt-Base, K.; Smith, R.; Tomchick, D. R.; Benning, M. M.; Winkelmann, D. A.; Wesenberg, G.; Holden, H. M. *Science* **1993**, *261*, 50–58.
139. Rayment, I. *Methods Enzymol.* **1997**, *276*, 171–179.
140. Rypniewski, W. R.; Holden, H. M.; Rayment, I. *Biochemistry* **1993**, *32*, 9851–9858.
141. Kurinov, I. V.; Mao, C.; Irvin, J. D.; Uckun, F. M. *Biochem. Biophys. Res. Commun.* **2000**, *275*, 549–552.
142. Schubot, F. D.; Waugh, D. S. *Acta Crystallogr.* **2004**, *D60*, 1981–1986.
143. Kristensen, O.; Laurberg, M.; Gajhede, M. *Acta Crystallogr.* **2002**, *D58*, 1198–1200.
144. Heras, B.; Martin, J. L. *Acta Crystallogr.* **2005**, *D61*, 1173–1180.
145. Kriminski, S.; Caylor, C. L.; Nonato, M. C.; Finkelstein, K. D.; Thorne, R. E. *Acta Crystallogr.* **2002**, *D58*, 459–471.
146. Juers, D. H.; Matthews, B. W. *Q. Rev. Biophys.* **2004**, *37*, 105–119.
147. Hope, H. *Acta Crystallogr.* **1988**, *B44*, 22–26.
148. Hope, H. *Annu. Rev. Biophys. Biophys. Chem.* **1990**, *19*, 107–126.
149. Young, A. C. M.; Dewan, J. C.; Nave, C.; Tilton, R. F. *J. Appl. Crystallogr.* **1993**, *26*, 309–319.
150. Rodgers, D. W. *Structure* **1994**, *2*, 1135–1140.
151. Rodgers, D. W. *Methods Enzymol.* **1997**, *276*, 183–203.
152. Watowich, S. J.; Skehel, J. J.; Wiley, D. C. *Acta Crystallogr.* **1995**, *D51*, 7–12.
153. Chayen, N. E.; Boggon, T. J.; Cassetta, A.; Deacon, A.; Gleichmann, T.; Habash, J.; Harrop, S. J.; Helliwell, J. R.; Nieh, Y. P.; Peterson, M. R. et al. *Q. Rev. Biophys.* **1996**, *29*, 227–278.
154. Garman, E. F.; Schneider, T. R. *J. Appl. Crystallogr.* **1997**, *30*, 211–237.
155. Garman, E. F.; Garman, E. F. *Acta Crystallogr.* **1999**, *D55*, 1641–1653.
156. Harp, J. M.; Timm, D. E.; Bunick, G. J. *Acta Crystallogr.* **1998**, *D54*, 622–628.
157. Harp, J. M.; Hanson, B. L.; Timm, D. E.; Bunick, G. J. *Acta Crystallogr.* **1999**, *D55*, 1329–1334.
158. Kiefersauer, R.; Than, M. E.; Dobbek, H.; Gremer, L.; Melero, M.; Strobl, S.; Dias, J. M.; Soulimane, T.; Huber, R. *J. Appl. Crystallogr.* **2000**, *33*, 1223–1230.
159. Abergel, C. *Acta Crystallogr.* **2004**, *D60*, 1413–1416.
160. Pflugrath, J. W. *Methods* **2004**, *34*, 415–423.
161. Garman, E. F.; Mitchell, E. P. *J. Appl. Crystallogr.* **1996**, *29*, 584–587.
162. Parkin, S.; Hope, H. *J. Appl. Crystallogr.* **1998**, *31*, 945–953.
163. Riboldi-Tünncliffe, A.; Hilgenfeld, R. *J. Appl. Crystallogr.* **1999**, *32*, 1003–1005.
164. Thomaneck, U. F.; Parak, F.; Mössbauer, R. L.; Formanek, H.; Schwager, P.; Hoppe, W. *Acta Crystallogr.* **1973**, *A29*, 263–265.
165. Urayama, P.; Phillips, G. N., Jr.; Gruner, S. M. *Structure* **2002**, *10*, 51–60.
166. Kim, C. U.; Kapfer, R.; Gruner, S. M. *Acta Crystallogr.* **2005**, *D61*, 881–890.
167. Rossmann, M. G. *Acta Crystallogr.* **1990**, *A46*, 73–82.
168. Storoni, L. C.; McCoy, A. J.; Read, R. J. *Acta Crystallogr.* **2004**, *D60*, 432–438.
169. McCoy, A. J.; Grosse-Kunstleve, R. W.; Storoni, L. C.; Read, R. J. *Acta Crystallogr.* **2005**, *D61*, 458–464.
170. Schwarzenbacher, R.; Godzik, A.; Grzechnik, S. K.; Jaroszewski, L. *Acta Crystallogr.* **2004**, *D60*, 1229–1236.
171. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235–242.
172. Berman, H. M.; Battistuz, T.; Bhat, T. N.; Bluhm, W. F.; Bourne, P. E.; Burkhardt, K.; Feng, Z.; Gilliland, G. L.; Iype, L.; Jain, S. et al. *Acta Crystallogr.* **2002**, *D58*, 899–907.
173. Hendrickson, W. A.; Ogata, C. M. *Methods Enzymol.* **1997**, *276*, 494–523.
174. Hendrickson, W. A.; Horton, J. R.; LeMaster, D. M. *EMBO J.* **1990**, *9*, 1665–1672.
175. Doublé, S. *Methods Enzymol.* **1997**, *276*, 523–530.
176. Smith, J. L.; Thompson, A. *Structure* **1998**, *6*, 815–819.
177. Blundell, T. L.; Johnson, L. N. In *Protein Crystallography*; Academic Press: London, UK, 1976, pp 183–239.
178. Petsko, G. A. *Methods Enzymol.* **1985**, *114*, 147–156.
179. Rould, M. A. *Methods Enzymol.* **1997**, *276*, 461–472.
180. Stura, E. A.; Gleichmann, T. In *Crystallization of Nucleic Acids and Proteins*; Ducruix, A., Giegé, R., Eds.; Oxford University Press: Oxford, UK, 1999, pp 365–390.
181. Cohen, S. L.; Chait, B. T. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 67–85.
182. Boggon, T. J.; Shapiro, L. *Structure* **2000**, *8*, R143–R149.
183. Gruen, M.; Becker, C.; Beste, A.; Reinstein, J.; Scheidig, A. J.; Goody, R. S. *Protein Sci.* **1999**, *8*, 2524–2528.
184. Buts, L.; Loris, R.; De Genst, E.; Oscarson, S.; Lahmann, M.; Messens, J.; Brosens, E.; Wyns, L.; De Greve, H.; Bouckaert, J. *Acta Crystallogr.* **2003**, *D59*, 1012–1015.
185. Cohen, A.; Ellis, P.; Kresge, N.; Soltis, S. M. *Acta Crystallogr.* **2001**, *D57*, 233–238.
186. Dauter, Z.; Dauter, M.; Rajashankar, K. R. *Acta Crystallogr.* **2000**, *D56*, 232–237.
187. Dauter, Z.; Dauter, M. *Structure* **2001**, *9*, R21–R26.
188. Nagem, R. A. P.; Dauter, Z.; Polikarpov, I. *Acta Crystallogr.* **2001**, *D57*, 996–1002.
189. Nagem, R. A. P.; Polikarpov, I.; Dauter, Z. *Methods Enzymol.* **2003**, *374*, 120–137.
190. Hendrickson, W. A.; Teeter, M. M. *Nature* **1981**, *290*, 107–113.
191. Weiss, M. S.; Sicker, T.; Djinic-Carugo, K.; Hilgenfeld, R. *Acta Crystallogr.* **2001**, *D57*, 689–695.
192. Weiss, M. S.; Sicker, T.; Hilgenfeld, R. *Structure* **2001**, *9*, 771–777.
193. Mueller-Dieckmann, C.; Polentarutti, M.; Djinic-Carugo, K.; Panjikar, S.; Tucker, P. A.; Weiss, M. S. *Acta Crystallogr.* **2004**, *D60*, 23–38.
194. Mueller-Dieckmann, C.; Panjikar, S.; Tucker, P. A.; Weiss, M. S. *Acta Crystallogr.* **2005**, *D61*, 1263–1272.
195. Matthews, B. W. *J. Mol. Biol.* **1968**, *33*, 491–497.
196. Card, G. L.; Blasdel, L.; England, B. P.; Zhang, C.; Suzuki, Y.; Gillette, S.; Fong, D.; Ibrahim, P. N.; Artis, D. R.; Bollag, G. et al. *Nat. Biotechnol.* **2005**, *23*, 201–207.
197. Erlanson, D. A.; McDowell, R. S.; O'Brien, T. *J. Med. Chem.* **2004**, *47*, 3463–3482.
198. Rees, D. C.; Congreve, M.; Murray, C. W.; Carr, R. *Nat. Rev. Drug Disc.* **2004**, *3*, 660–672.
199. Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. *Science* **1996**, *274*, 1531–1534.

200. Nienaber, V. L.; Richardson, P. L.; Klighofer, V.; Bouska, J. J.; Giranda, V. L.; Greer, J. *Nat. Biotechnol.* **2000**, *18*, 1105–1108.
201. Blundell, T. L.; Jhoti, H.; Abell, C. *Nat. Rev. Drug Disc.* **2002**, *1*, 45–54.
202. Blundell, T. L.; Patel, S. *Curr. Opin. Pharmacol.* **2004**, *4*, 490–496.
203. Hartshorn, M. J.; Murray, C. W.; Cleasby, A.; Fredrickson, M.; Tickle, I. J.; Jhoti, H. *J. Med. Chem.* **2005**, *48*, 403–413.
204. Congreve, M.; Carr, R.; Murray, C.; Jhoti, H. *Drug. Disc. Today* **2003**, *8*, 876–877.
205. Nienaber, V. L.; Davidson, D.; Edalji, R.; Giranda, V. L.; Klighofer, V.; Henkin, J.; Magdalinos, P.; Mantei, R.; Merrick, S.; Severin, J. M. et al. *Structure* **2000**, *8*, 553–563.
206. Hardy, J. A.; Wells, J. A. *Curr. Opin. Struct. Biol.* **2004**, *14*, 706–715.
207. Erlanson, D. A.; Hansen, S. K. *Curr. Opin. Chem. Biol.* **2004**, *8*, 399–406.
208. Erlanson, D. A.; Wells, J. A.; Braisted, A. C. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33*, 199–223.
209. Hardy, J. A.; Lam, J.; Nguyen, J. T.; O'Brien, T.; Wells, J. A. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 12461–12466.
210. O'Brien, T.; Fahr, B. T.; Sopko, M. M.; Lam, J. W.; Waal, N. D.; Raimundo, B. C.; Purkey, H. E.; Pham, P.; Romanowski, M. J. *Acta Crystallogr.* **2005**, *F61*, 451–458.
211. Lesuisse, D.; Lange, G.; Deprez, P.; Bénard, D.; Schoot, B.; Delettre, G.; Marquette, J. P.; Broto, P.; Jean-Baptiste, V.; Bichet, P. et al. *J. Med. Chem.* **2002**, *45*, 2379–2387.
212. Ramstrom, O.; Lehn, J. M. *Nat. Rev. Drug Disc.* **2002**, *1*, 26–36.
213. Congreve, M. S.; Davis, D. J.; Devine, L.; Granata, C.; O'Reilly, M.; Wyatt, P. G.; Jhoti, H. *Angew. Chem., Int. Ed. Engl.* **2003**, *42*, 4479–4482.
214. Burley, S. K. *Nat. Struct. Biol.* **2000**, *7*, 932–934.
215. Stevens, R. C.; Yokoyama, S.; Wilson, I. A. *Science* **2001**, *294*, 89–92.
216. Bussow, K.; Nordhoff, E.; Lubbert, C.; Lehrach, H.; Walter, G. *Genomics* **2000**, *65*, 1–8.
217. Christendat, D.; Yee, A.; Dharamsi, A.; Kluger, Y.; Savchenko, A.; Cort, J. R.; Booth, V.; Mackereth, C. D.; Saridakis, V.; Ekiel, I. et al. *Nat. Struct. Biol.* **2000**, *7*, 903–909.
218. Nasoff, M.; Bergseid, M.; Hoeffler, J. P.; Heyman, J. A. *Methods Enzymol.* **2000**, *328*, 515–529.
219. Knaust, R. K.; Nordlund, P. *Anal. Biochem.* **2001**, *297*, 79–85.
220. Dobrovetsky, E.; Lu, M. L.; Andorn-Broza, R.; Khutoreskaya, G.; Bray, J. E.; Savchenko, A.; Arrowsmith, C. H.; Edwards, A. M.; Koth, C. M. *J. Struct. Funct. Genomics* **2005**, *6*, 33–50.
221. Kigawa, T.; Yabuki, T.; Yoshida, Y.; Tsutsui, M.; Ito, Y.; Shibata, T.; Yokoyama, S. *FEBS Lett.* **1999**, *442*, 15–19.
222. Kigawa, T.; Yabuki, T.; Matsuda, N.; Matsuda, T.; Nakajima, R.; Tanaka, A.; Yokoyama, S. *J. Struct. Funct. Genomics* **2004**, *5*, 6368.
223. Kuhn, P.; Wilson, K.; Patch, M. G.; Stevens, R. C. *Curr. Opin. Chem. Biol.* **2002**, *6*, 704–710.
224. Hosfield, D.; Palan, J.; Hilgers, M.; Scheibe, D.; McRee, D. E.; Stevens, R. C. *J. Struct. Biol.* **2003**, *142*, 207–217.
225. Weselak, M.; Patch, M. G.; Selby, T. L.; Knebel, G.; Stevens, R. C. *Methods Enzymol.* **2003**, *368*, 45–76.
226. Cohen, A. E.; Ellis, P. J.; Miller, M. D.; Deacon, A. M.; Phizackerley, R. P. *J. Appl. Crystallogr.* **2002**, *35*, 720–726.
227. Karain, W. I.; Bourenkov, G. P.; Blume, H.; Bartunik, H. D. *Acta Crystallogr.* **2002**, *D58*, 1519–1522.
228. Pohl, E.; Ristau, U.; Gehrmann, T.; Jahn, D.; Robrahn, B.; Malthan, D.; Döbler, H.; Hermes, C. *J. Synchrotron Radiat.* **2004**, *11*, 372–377.
229. Snell, G.; Cork, C.; Nordmeyer, R.; Cornell, E.; Meigs, G.; Yegian, D.; Jaklevic, J.; Jin, J.; Stevens, R. C.; Earnest, T. *Structure* **2004**, *12*, 537–545.
230. Muchmore, S. W.; Olson, J.; Jones, R.; Pan, J.; Blum, M.; Greer, J.; Merrick, S. M.; Magdalinos, P.; Nienaber, V. L. *Structure* **2000**, *8*, R243–R246.
231. Abad-Zapatero, C. *Acta Crystallogr.* **2005**, *D61*, 1432–1435.
232. Turk, D. Towards Automatic Macromolecular Crystal Structure Determination. In *Methods in Macromolecular Crystallography*; Turk, D., Johnson, L., Eds.; NATO Science Series I, Vol. 325; IOS Press: Amsterdam, Netherlands, 2001; pp 148–155.
233. Adams, P. D.; Gopal, K.; Grosse-Kunstleve, R. W.; Hung, L.; Ioerger, T. R.; McCoy, A. J.; Moriarty, N. W.; Pai, R. K.; Read, R. J.; Romo, T. D. et al. *J. Synchrotron Radiat.* **2004**, *11*, 53–55.
234. Kroemer, M.; Dreyer, M. K.; Wéndt, K. U. *Acta Crystallogr.* **2004**, *D60*, 1679–1682.
235. Ness, S. R.; de Graaff, R. A.; Abrahams, J. P.; Pannu, N. S. *Structure* **2004**, *12*, 1753–1761.
236. Holton, J.; Alber, T. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 1537–1542.
237. Panjikar, S.; Parthasarathy, V.; Lamzin, V. S.; Weiss, M. S.; Tucker, P. A. *Acta Crystallogr.* **2005**, *D61*, 449–457.
238. Schüttelkopf, A. W.; van Aalten, D. M. F. *Acta Crystallogr.* **2004**, *D60*, 1355–1363.
239. Diller, D. J.; Pohl, E.; Redinbo, M. R.; Hovey, B. T.; Hol, W. G. *Proteins* **1999**, *36*, 512–525.
240. Oldfield, T. J. *Acta Crystallogr.* **2001**, *D57*, 696–705.
241. Oldfield, T. *Methods Enzymol.* **2003**, *374*, 271–300.
242. Zwart, P. H.; Langer, G. G.; Lamzin, V. S. *Acta Crystallogr.* **2004**, *D60*, 2230–2239.
243. Aishima, J.; Russel, D. S.; Guibas, L. J.; Adams, P. D.; Brunger, A. T. *Acta Crystallogr.* **2005**, *D61*, 1354–1363.

Biographies



Tanis Hogg was born in Saskatoon, Canada, and spent his formative years in the town of Kamloops, where he developed an early interest in the fields of biology and chemistry. He studied general sciences at Thompson Rivers University (Kamloops) and later transferred to the University of British Columbia (Vancouver), where he earned his BSc in biochemistry in 1993. After graduation, he took a temporary position with Baxter Pharmaceuticals in Burnaby, which stimulated his interest to pursue graduate studies in the field of protein crystallography and structure-based drug design. He moved to Jena, Germany, at the end of 1996 and studied under the supervision of Prof Rolf Hilgenfeld at the Institute of Molecular Biotechnology. There, he carried out crystallographic studies on ribosome-associated GTP-binding proteins including the bacterial drug target elongation factor Tu, and was awarded his PhD in 2001. After receiving his PhD, he became group leader in crystallography and structure-based drug design at JenaDrugDiscovery, GmbH. In 2004, he accepted a position at the Institute of Biochemistry, University of Lübeck, Germany, where his group is focused on macromolecular crystallography and rational design of novel antibiotics targeting pathogenic bacteria.



Rolf Hilgenfeld studied chemistry at the universities of Göttingen and Freiburg, Germany. He did his PhD in protein crystallography at the Free University of Berlin and after a postdoctoral stay at the Biocenter of the University of Basel, Switzerland, he joined Hoechst AG, the pharmaceutical company in Frankfurt, to build a macromolecular crystallography laboratory. During his 9 years in the company, he and his colleagues worked on the design of new insulin variants with the goal of improving the pharmacokinetics of the hormone. A major achievement of these efforts was the creation of a long-acting insulin, which has been introduced into the market as 'Lantus'. Rolf Hilgenfeld was also among the first scientists to determine the structure of the HIV-1 protease and to design inhibitors against this target. He also elucidated the structure of elongation factor Tu and studied its interaction with antibiotics. In 1995, he moved to the University of Jena to take over the Chair of Structural Biochemistry, in combination with the position of Head of the Crystallography Department at the newly founded Institute of Molecular Biotechnology. Since 2003, Rolf Hilgenfeld has been Full Professor of Biochemistry at the University of Lübeck, Germany. Today, his research focuses

on the molecular basis of infectious diseases by bacteria such as *Legionella pneumophila* and *Chlamydia* and by RNA viruses. During the global SARS epidemic of 2003, he published the crystal structure of the coronavirus main proteinase and proposed a first inhibitor against the disease. His Lübeck laboratory follows an integrated approach to drug discovery against infectious agents, which includes comparative proteomics, molecular biology, x-ray crystallography, drug design, and chemical synthesis of inhibitors.